Research Note 83-8

AD A127359

USER-COMPUTER INTERACTIONS: SOME PROBLEMS FOR HUMAN FACTORS RESEARCH

R.S. Nickerson, T.H. Meyer, D.C. Miller,
and R.W. Pew

BOLT BERANEK AND NEWMAN INC.

BASIC RESEARCH

U. S. Army

Research Institute for the Behavioral and Social Sciences

September 1981

83 04 21 121

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Note 83-8	2. GOVT ACCESSION NO. AD-A127089	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) User-Computer Interaction: Some Problems for Human Factors Research		5. TYPE OF REPORT & PERIOD COVERED Final Report 17 July 80 -- 16 July 81
		6. PERFORMING ORG. REPORT NUMBER BBN Report No. 4719
7. AUTHOR(s) R.S. Nickerson, T.H. Meyer, D.C. Miller, and R.W. Pew (BBN Inc.)		8. CONTRACT OR GRANT NUMBER(s) MDA 903-80-C-0551
9. PERFORMING ORGANIZATION NAME AND ADDRESS Bolt Beranek and Newman Inc. 50 Moulton Street Cambridge, MA 02238		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS Army Research Institute for the Behavioral and Social Sciences. 5001 Eisenhower Avenue, Alexandria, Virginia 22333		12. REPORT DATE September 1981
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

User-computer interaction
Human Factors Research
Communication
User-computer dialogue
Information systems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report discusses a variety of research problems that relate to the use of interactive computer systems in military contexts. It begins with a review of several documents that describe planned or anticipated changes in the deployment and use of computers by the Department of Defense in the near-term future. It then discusses several generic military functions that involve the use of computer systems. Research problems are discussed under the general topics of user issues, interface issues, and system issues. Finally several methodo-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) item #20 - continued

logical issues relating to human factors research on computer-based systems are considered.

A

TABLE OF CONTENTS

		Page
1.	INTRODUCTION AND BACKGROUND	5
1.1	Purpose and Scope of Study	5
1.2	The Need for Human Factors Research	6
1.3	Approach	10
1.4	Organization of Report	11
1.5	The Background of Technological Change	14
1.5.1	Recent Trends	15
1.5.2	Near-term Expectations	17
1.6	Some Anticipated Changes in DoD Computer Systems	25
1.6.1	The WWMCCS Information System and Its Planned Modernization	26
1.6.2	CENTACS and the Military Computer Family	34
1.6.3	DARCOM's Blueprint for Information Processing in the 1980s	38
1.6.4	Report of Advisory Committee on Information Structure and Functions for the EOP	41
1.6.5	Report of the IDA on C^3I Data Communication Networks	43
1.6.6	Summary Comment	45
2.	MILITARY APPLICATIONS OF INTERACTIVE COMPUTER SYSTEMS	49
2.1	Communication	49
2.1.1	Message Processing	51
2.1.2	Content-dependent Message Routing	54
2.1.3	Teleconferencing	55
2.1.4	Mobile Digital Communications (Packet Radio)	59
2.1.5	Foot Patrol Communication	60
2.1.6	Reporting of Location Information	62
2.1.7	The Problem of Communication Gaps or Failures	62
2.1.8	Communication Security	64
2.1.9	Communication System Criteria	65

2.2 Situation Assessment 65
2.3 Office Automation and Information Management 66
2.4 Monitoring and Supervisory Control 67

3. USER ISSUES 71

3.1 Person-Computer Function Allocation 71
3.2 Types of Users 75
3.3 User "Styles" 78
3.4 Psychological Barriers to Computer Use 78
3.5 User Acceptance of Innovation 80
3.6 Determination of Job Requirements 81
3.7 Skill Maintenance 81
3.8 Safeguards Against User Error 82
3.9 Organizational Impact of Computer-based Systems 85

4. INTERFACE ISSUES 87

4.1 Displays 93
4.1.1 A Note on Display Hardware 94
4.1.2 Coding Parameters for Dynamic Displays 96
4.1.3 Information Access 99
4.1.4 Dynamic Displays 105
4.1.5 Content-format Issues Regarding Information I/O 106
4.1.6 Computer-controlled Maps 107
4.2 Speech 108
4.3 Other High-bandwidth Input Methods 110
4.4 User-Computer Dialogue 111
4.4.1 Menus 112
4.4.2 Formatted Inputs 113
4.4.3 Question-and-Answer Inputs 113
4.4.4 Limited-Syntax Command Languages 114
4.4.5 Special-Purpose Function Keyboards 116
4.4.6 Natural Language 118
4.4.7 Graphics Language Development 121
4.4.8 Customizable Interfaces 122
4.5 Tools and Procedures 122
4.5.1 System Design and User Models 122
4.5.2 Complexity 129
4.5.3 Documentation and On-line User Aids 134
4.5.4 Information Finding Techniques 134
4.5.5 Knowledge-based Dialogue Tools 139

5. SYSTEM ISSUES 143

5.1 System Architecture 143
5.2 Survivability 147
5.3 Interoperability, Inter-system Compatibility 154
5.4 Security 163
5.5 Responsiveness 167
5.5.1 Basic Capacity 168
5.5.2 Responsiveness and the Psychology of Acceptance 169
5.5.3 Architecture 172

6. METHODOLOGICAL ISSUES 177

6.1 Major Approaches to the Study of User-computer Interaction 177
6.1.1 Observation 177
6.1.2 Controlled Experimentation 178
6.1.3 Modelling and Simulation 179
6.2 Some Other Methodological Issues 185
6.2.1 Cognitive Workload Measurement 185
6.2.2 On Identifying What is Wrong with Existing Systems 185
6.2.3 Software Evaluation 187
6.2.4 Methodology for Computer System Design 189

7. REFERENCES 193

APPENDIX A. PLANNED CAPABILITIES OF FOUR OPERATIONAL FAMILIES OF FUNCTIONS FOR THE WWMCCS INFORMATION SYSTEM 199

ABSTRACT

This report discusses a variety of research problems that relate to the use of interactive computer systems in military contexts. It begins with a review of several documents that describe planned or anticipated changes in the deployment and use of computers by the Department of Defense in the near-term future. It then discusses several generic military functions that involve the use of computer systems. Research problems are discussed under the general topics of user issues, interface issues, and system issues. Finally, several methodological issues relating to human factors research on computer-based systems are considered.

ACKNOWLEDGMENTS

In addition to the authors of this report, major contributors to the project were Sheldon Baron, Jerry Burchfiel, and Charlotte Mooers. Participants in the two BBN workshops mentioned in Section 1.3 were as follows:

Workshop on Tactical Communications:

Michael Beeler
Carl Feehrer
Charlotte Mooers
Duncan Miller
Raymond Nickerson
Conrad Nuthmann
Richard Pew
Oliver Selfridge
Albert Stevens
Robert Thomas
Jerry Wolf

Workshop on Issues in Interactive System Design:

Debra Deutsch
Norton Greenfeld
Dan Massey
Duncan Miller
Theodore Myer
Raymond Nickerson
Richard Pew
Candace Sidner
Albert Stevens
Janet Walker

We are also indebted to Ray Sidorsky of ARI for arranging visits to Ft. Monmouth and Ft. Leavenworth and to the various people at these installations who served as hosts during our

visits or briefed us on their activities. These included Alan Wood at Ft. Monmouth and Robert Andrews at Ft. Leavenworth.

1. INTRODUCTION AND BACKGROUND

1.1 Purpose and Scope of Study

The purpose of this report is to identify some research problems relating to user-computer interaction that appear to have relevance for military operations in the next few decades. More specifically, our intent is to focus on system-independent research issues that relate to human factors aspects of military operations and that might be appropriate topics for Army-funded research. It is probably important to be clear, therefore, about what we mean by "system-independent" research.

For purposes of this report, system-independent research will be conceived sufficiently broadly to include research that is addressed to questions that have obvious practical implications, but not sufficiently broadly to include questions relating to the design of specific pieces of hardware, software or systems. Our interest will be primarily in identifying generic problems that will yield system-independent solutions with applicability to many systems functioning in a wide variety of situations. We believe the identification of such problems is a continuing challenge for human factors researchers and that it is a goal worth pursuing. A danger of research that is addressed to questions of the merits or limitations of specific systems is that the results will fail to generalize and therefore be of little use to anybody other than the developers of those particular systems, and perhaps not even to them. Inasmuch as research progress tends to be slow relative to system development cycles, research that is narrowly focused is almost invariably doomed to produce results that are obsolete by the time they are available for use.

1.2 The Need for Human Factors Research

The need for research on human factors aspects of the design and operation of computer-based systems derives from a number of facts, among which are the following:

- Increasing complexity of the computer-based systems that are being developed: As computing resources decrease in cost, they are being applied to an ever expanding range of problems, and the systems that are being developed are being given greater and greater functionality. In many cases the existing systems are sufficiently complex that their users have a very poor conceptual understanding of how they operate or what they can or cannot be expected to do. People learn enough to use such systems in routine ways but often are at a total loss to know how to deal with system failures or malfunctions. As systems continue to increase in complexity, the mismatch between their capabilities and the level of understanding of the average user will continue to increase as well. How to cope with this fact is only one of the several manifestations of the problem of dealing with complexity that the continuing evolution of information technology presents. Moreover, it is a human factors problem inasmuch as the difficulty lies not with the complexity of systems *per se* but with the challenge that this complexity poses for their maintenance and operation by human beings.

- Consequences of failure due to human error: A corollary to increased complexity of tools typically is an increase in the seriousness of potential consequences of user error. While a power saw is a more effective tool than a hand saw for most applications, it also is a considerably more dangerous one. A 747 jet aircraft is a far more powerful and complex means of transporting people from place to place than is a two passenger aircraft, and the potential consequences of pilot error are clearly much greater in the former case than in the latter. A computer-based command, control, and communication system that provides a high-level decision maker with timely information needed to assess military threat and the ability to communicate action orders quickly to widely dispersed elements can represent a considerable advantage to a user by decreasing the time required to react decisively to a developing situation; the consequences of inappropriate or ill-advised action, however, can obviously be grave.

- Shortage of highly-skilled personnel: Much concern has been expressed in recent years that the termination of the draft has resulted in a dramatic decrease in the education and skill level of the average military recruit. This has implications both for the importance of human factors in the design of military systems and for training. The coupling of increasing complexity and sophistication of systems with decreasing qualifications of the personnel to operate them is a disturbing trend. Research focused on the objective of identifying ways of making complex systems less subject to human error and on more effective methods for operator training are not the only approaches that will have to be taken to resolve this problem, but they clearly both are necessary ones.

- Limited time for training and high turnover of military personnel: The problem mentioned above is compounded by the fact that fewer people are making a career of the military, so large investments in training are less cost effective than they otherwise would be. Moreover, there is a paradox associated with effective training programs. The more successful a training program is in giving an individual high-level skills, the easier it is for that individual to find attractive job opportunities outside the military. This again points out the importance of finding ways to design systems so they can be used effectively in spite of the limited qualifications and specialized training of available personnel.

Initial efforts to apply human factors principles to interactive computer systems focused on translating and applying to this new field the existing knowledge that had been developed in other contexts. Areas in which a substantial body of knowledge already existed include

- Information coding dimensions, including both alphanumerics (character size and font, mnemonics, abbreviations, etc.) and symbols (shape, size, color, blink rates, etc.).
- Readability and intelligibility of instructions and warnings.
- Display characteristics, including character size, font, and spacing; phosphor characteristics, refresh rates, etc. (although some of these issues had to be readdressed in the light of limitations produced by dot matrix character generation methods).
- Human motor performance, reaction times, and information processing rates.
- Keyboard layout, physical properties, and operating procedures.
- Characteristics of some other potential input control devices, such as joysticks and trackballs, which were already in use in other applications.

The existing knowledge in some of these areas was more or less directly applicable; in other areas, the issues involved in interactive systems had few precedents, and new studies were essential. Timing issues in interactive dialogues represent a clear example from the latter category.

Recently, efforts have begun toward enlarging the focus of analyses of human factors in computer systems to encompass the entire system environment with which a user may interact. There is very little formal experimentation on which to draw, however. Ramsey and Atwood (1979) characterize the situation as follows:

> In some well established research areas, such as keyboard design, and certain physical properties of displays, guidelines exist which are reasonably good and fairly detailed. Such guidelines may be quite helpful in the design of a console or other interface device for a system, or even in the selection of an appropriate off-the-shelf input/output device. As we progress toward the more central issues in interactive systems, such as their basic informational properties, user aids, and dialogue methods, available guidelines become sketchy and eventually nonexistent. The interactive system designer is given little human factors guidance with respect to the most basic design decisions. In fact, the areas in which existing guidelines concentrate are often not even under the control of the designer, who may have more freedom with respect to dialogue and problem-solving aids than with respect to terminal design or selection (p. 2).

Most of the design guidelines that are beginning to emerge are based on the application of systems analysis techniques in the context of specific interactive systems. As certain systems gain recognition for having well-designed user interfaces, their salutary characteristics can be incorporated into guidelines that can be extrapolated for use in the design of other interactive systems. There remains, however, the need not only to organize what has already been discovered about how to design effective interfaces, but also to develop, through research, a new knowledge base that can help assure both the usefulness and the useability of new systems whose functionality will extend beyond any that currently exist.

1.3 Approach

In addition to our own experience in conducting research and development on interactive computer systems, the ideas in this report came from three sources: (1) literature on existing and planned military systems and on research on person-computer interaction, (2) visits to two Army sites, and (3) two one-day workshops at Bolt Beranek and Newman.

The Army sites that were visited were Ft. Monmouth and Ft. Leavenworth. At each installation, we talked with several people representing various units responsible for the development and/or operation of computer-based systems. In each case we also were provided with some documentation that has been referenced in this report.

The two workshops that were held at BBN focused on "Tactical Communications" and "Issues in Interactive System Design." At each of these workshops, we had approximately ten participants, each of whom had research or operational experience relevant to the workshop topic. These workshops were run more or less in the style of brainstorming sessions, at which people were encouraged to put forward ideas regarding researchable problems that were appropriate for this project. Not surprisingly, not all of the ideas that emerged found their way into this report; however, many of them did.

The hosts for the site visits and participants in BBN workshops are listed in the preface to this report.

1.4 Organization of Report

In the remainder of Section 1 of this report we provide a frame of reference for what follows by describing some relevant developments in information technology in general and in military information systems in particular. In Section 2 we discuss several categories of applications of interactive computer systems in various military contexts, and identify a number of unresolved issues relating to these applications.

In Sections 3, 4, and 5 we attempt to identify some generic problems involving the effectiveness with which a human being will perform as part of a user-computer system. The discussion considers both user and system issues in some detail.

In Section 3 we focus on the human side of the equation. The discussion first explores the psychological, educational, and job structure issues that must be addressed in ensuring an effective user-computer interaction, and then addresses additional concerns that arise when several individuals interact with a computer system in ways that may affect organizational structure.

In Section 4 we address issues relating to the user-system interface, which is typically considered to be the major point at which human factors issues involving user-computer interaction arise.

In Section 5 we consider aspects of system design ordinarily considered to be "below" the level of the user-computer interface. The discussion begins with the functions and tools provided by a system and the organization of information within that system. We then consider system architecture -- the major hardware and software components of a system and how they are interconnected.

Taken together, Sections 2 through 5 suggest numerous research topics, but say little about the methodology for attacking them. In Section 6 we address some methodological issues relating to the performance of human factors research on interactive systems. A key focus of the discussion is the difficulty arising from the complexity of such systems, particularly those that support organized groups of users. The discussion considers techniques for observing, modeling, and designing interactive systems.

This organization, like most any other that one might impose on the material in this report, is arbitrary to a degree. The material does not lend itself to partitioning into nonoverlapping categories. While the distinction among users issues, interface issues, and system issues, for example, is conceptually meaningful, it is not possible to say very much about any one of these topics without commenting also on the others. The headings of the major sections are indicative of the primary focus of those sections; however, the overlap among sections is fairly substantial and necessarily so.

Although our focus is primarily on research issues that relate to user-computer interaction in military contexts, insofar as the problems that are addressed are relevant also to other-than-military contexts, we discuss them in general terms. In fact, given the plans for uses of computer systems by the military, as described in documents discussed in the remainder of Section 1, most civilian applications of computers have their analogs in military environments. Even such functions as those associated with the concept of office automation, which one typically might not relate to military (and especially tactical) environments, are expected to be very important in systems such as the WWMCCS system as presented in its modernization plan.

1.5 The Background of Technological Change

This study is motivated by rapid technological change. Without such change, we would not now have a problem of user-computer interaction. Were the era of change to have ended, we would have a static situation that could be attacked in reasonably relaxed fashion. However, the facts are that there has been rapid change for three decades, the rate of change is increasing, and there is no clear end in sight. Since it is against this background that the study proceeds, it seems reasonable to make some comments about the background itself.

In this section, we briefly review the recent past of information technology and outline what seem to be clear directions for its development in the near future. At this stage, the coverage is quite general, although we do link the major developments to systems and applications of present or potential use in the Army.

The material that follows is adapted from a chapter prepared by one of the authors of this report (Nickerson, in press a) for a forthcoming book on anticipated developments in information technology and their psychological and social implications (Kasschau, Lachman & Laughery, in press).

1.5.1 Recent Trends

An examination of the recent past of computer technology reveals some major trends: decreasing component costs, decreasing size and increasing packing density of components, decreasing power requirements, increasing speed, increasing reliability, increasing size of market, and increasing distribution of and accessibility to computing resources. The following observations serve to illustrate the speed with which the technology is moving.

- The cost of production of a logic gate has gone from about $10 in 1960 to about 10 cents in 1970 to less than one cent in 1980.
- The cost of dynamic random-access memory (RAM) has gone from about one cent per bit in 1970 to about .05 cent per bit in 1980.
- The size of the smallest feature on an integrated circuit has gone from about 10 microns in 1970 to about three microns in 1980.
- The number of active element groups that can be placed on a single semiconductor chip has gone from less than 10 in 1960 to a few thousand in 1970 to approximately 70,000 in 1980.* The rate of increase has been roughly an order of magnitude every five years since 1960.
- Random-access memory power dissipation has gone from

---

*Since this material was written, Bell Laboratories has developed a microprocessor (the MAC-32) that has over 100,000 components on a single chip, and Hewlett-Packard has developed one with over 450,000 (Johnson, 1981).

about 500 microwatts per bit in 1970 to about four microwatts per bit in 1980.

- Access time for dynamic RAM has gone from about 400 nanoseconds in 1970 to about 150 nanoseconds in 1980.
- The reliability of logic gates has increased by about 5 orders of magnitude in two decades.
- In 1979 microprocessor sales increased by more than 35% over 1978. At about the same time the cost of a micro processor dropped from about $65 to about $5 over a period of 18 months.
- In spite of enormous efforts to expand production capacity -- $800 million for new plant and equipment by the semiconductor industry in 1979 -- the demand appears to be growing faster than the supply.
- Lead times for many semiconductor components are now six months to one year.
- The estimated number of active element groups in the average U.S. home has gone from about 10 in 1940 to about 100 in 1960 to a few thousand in 1980.

Collectively, these trends represent an enormous increase in our ability to manipulate, store, and transmit very large amounts of information very rapidly and at steadily decreasing cost. If one wants to capture in a single term what has really increased, one could do worse than the term "access bandwidth." Thanks to Gutenberg and the later discovery of how to make paper from linen rags, it has been possible for some time to store large amounts of information in certain locations, such as the major libraries of the world. In our own century, the development of electronic means of storing and accessing information has increased the ease of both storage and retrieval. What current developments in information technology are doing is making it increasingly feasible to store truly huge amounts of information very economically, to make this information far more immediately accessible to people who want to use it, and not only to deliver to the users the information in prepackaged form but also to provide them with processes and procedures for manipulating it in useful ways.

1.5.2 Near-term Expectations

Most, if not all, of the trends noted above will undoubtedly continue into the foreseeable future. The costs of computer components will continue to decrease, as will their size and power requirements. Speed will increase, as will reliability. The demand for components and systems will continue to grow. Applications will continue to proliferate.

These predictions all seem reasonably safe. Of course, they amount to nothing more than a timidly qualitative projection of the past. When one attempts to quantify what will happen one begins to run the risk of almost certainly being wrong. Nevertheless, one can find some guesses of a quantitative sort regarding what will happen during the next few years. Bylinsky (1981), for example, has reported a prediction that the production cost of integrated circuit memory will drop to less than .005 cent per bit by the middle of the decade. The size of the smallest feature on an integrated circuit is expected to decrease to less than 1 micron by the mid-1980's (Kahn, 1978). Young (1981) anticipates several million components on a VLSI chip by the end of the decade. Leonard (1980) has predicted that the worldwide demand for semiconductor random-access memory will be about 20 trillion bits by 1982 (up from about 1.5 trillion in 1980). The number of active element groups in the average U.S. home is expected to be close to .5 million by the end of the decade, up from a few thousand in 1980 (Robinson, 1980).

Phipps (in press) points out that 15 years ago, the per capita consumption of electronic circuits in the United States was approximately three, whereas today it is roughly 10,000, and by 1990 it is expected to be about 2 million. This means that by the end of this decade the number of electronic circuits available for use in the United States will be roughly 2 million times the size of the population, a prediction that is not only believable, but possibly even conservative. However, it is difficult to imagine what its realization will mean.

Some of the expected advances can occur as the result of further refinement of existing techniques. Others will require the development of qualitatively different ways of doing things. Production of integrated circuits with submicron feature sizes, to illustrate the latter case, will require the use of X-rays, electron beams, or some other form of relatively short wavelength radiation, inasmuch as the current feature size is close to the limit imposed by the resolving power of visible light.

Two developments that are currently causing considerable excitement in the computer industry, and that are likely to play a significant role in advancing information technology in the 1980s and beyond, are the magnetic bubble memory and the Josephson junction. In a magnetic-bubble memory, one bit of information is represented by the presence or absence of a tiny area (magnetic bubble) that has a direction of polarization opposite to that of bulk of the material in which it is embedded. The most common material for bubble memories at the present is synthetic garnet, which permits the use of bubbles about .5 micrometers in diameter. Bubble memories built with thin films of metallic glasses may make feasible the use of bubbles of about .1 micrometer in diameter, thus allowing a 25-fold increase in packing density over the current state of the art. Normally-conducting vortexes in superconducting metallic glasses offer the possibility of another two-orders-of-magnitude increase in storage density over the .1 micrometer bubble memories, inasmuch as the vortexes measure only .005 to .01 micrometer (Chaudhari, Giessen & Turnbull, 1980).

The Josephson junction, a device invented by Brian Josephson in 1962, could replace the transistor as the fundamental element of computer technology, just as the transistor replaced the vacuum tube. The Josephson junction works on principles different from those of either the vacuum tube or the transistor, but like both of them it can act as a switch for an electronic signal. Among the advantages of the junction are the speed with which it can switch from one state to another (about six picoseconds) and its relatively small power requirements (because it is a superconducting device). It also can store information as well as function as a switch. The expectation is that when Josephson junction technology is further developed, it will be possible to construct computers from these devices that will consume a fraction of the power of today's microcomputers and will have memory cycle times of less than one nanosecond (Matiso, 1980).

Thus even with techniques that are already known, albeit in some cases still at an experimental stage, one can see a continuation at least for the near future of the recent trends in size, speed, power requirements and so on. That is not to say that these trends can continue indefinitely. Although the most exciting developments are likely to come from the least predictable quarters, there are -- the physicists tell us -- some fundamental limits that some of these trends will sooner or later encounter. Size can be reduced only so far, for example, until it runs into the limits of atomic structure. Indeed, the .01 micrometer mentioned in connection with bubble memories is within two orders of magnitude of the diameter of a hydrogen atom. Moreover, problems arise long before such fundamental limits are approached.

While the question of how far current trends can continue before encountering fundamental limitations is an open one, it is not clear that it has any very significant immediate practical implications. Moreover, it is likely that before current trends are pressed to their limits, fundamentally new computing architectures and approaches will be developed that will make these limits irrelevant. What is clear is that the computing resources that will be available in the foreseeable future will be enormous and widely available. Long before progress in information technology is halted because of fundamental physical limitations, we are likely to encounter obstacles of a quite different kind, namely our limited ability to exploit effectively the potential that the technology represents.

What kinds of applications of information technology can we anticipate in the foreseeable future? If we assume that computer resources are going to be increasingly widely distributed, readily accessible and inexpensive, then efforts to predict how they will be used are probably doomed to failure, or at least to gross understatement.

However, one can list a number of ways in which it now seems likely that information technology will be developed and used in the next few years. The following list of possibilities includes general techniques for information handling and broad application areas useful to both military and civilian sectors.

- Development of new architectures for processors and memory utilizing new materials and techniques, including gallium arsenide logic, superconducting devices such as the Josephson junction, charge-coupled devices, magnetic bubbles, optical communication and storage, and three-dimensional integrated circuits.

- Increasing use of fiber optics, microwave, and satellite technologies in communication systems, with the effect of broadening bandwidth greatly and providing increased accessibility to information of almost every sort.

- Satellite transmission directly to information users, increasing enormously the amount of information that can be delivered.

- Development and refinement of software tools to help designers and programmers cope with the increasing complexity of their tasks.

- Increasing emphasis on distributed computing systems and parallel processing approaches to the solution of complex problems.

- Much greater attention to practical applications of artificial intelligence, made feasible by the availability of sufficient computing speed and memory capacity provided by very-large-scale integration.

- Increasing use of speech as a means of communication between people and computers, and between people and people via computer networks.

- Widespread use of electronic mail, electronic funds transfer, and computer-mediated communication more generally.

- Increasingly powerful word-processing, document-preparation and information-management tools in the office.

- Increasing use of electronic means of composing, proofreading, editing, and disseminating "publications."

- Computer-based recruiting, cataloguing, and requisitioning; and proliferation of computer-mediated information services generally.

- Increasing involvement of computers in monitoring and control processes across a wide range of activities or systems: from the computerized fuel injection system in vehicles that optimizes energy usage by adapting the fuel mix to match the conditions of the moment, to the operation of a complete economy.

- Increasing use of automation and robotics.

- Increasing use of smart (and instructable) devices in offices and vehicles.

- Concentration of very large amounts of information of various types.

- Inexpensive ways of storing large amounts of information electronically in offices and other work places.

- Electronic accessibility from the work place or home of information stored in major repositories.

- Two-way real-time communication between broadcasting facilities and recipients.

Beyond these rather general developments, we can expect to see an increasingly important role of information technology in Army and other military systems, including its infusion into existing systems and the creation of new kinds of systems that could not exist without it. Examples include:

- Guidance systems for individual missiles.

- Air surveillance and attack warning systems.

- Tactical fire control systems.
- Battlefield information systems.
- Message communication systems.
- Distributed command and control systems.

A change that we are beginning to see, and that will become increasingly apparent, is a change in the predominant mode of computer use. The technique of time-sharing was developed in the early 1960s and has served us well through the last two decades. There will probably continue to be a demand for time-sharing services throughout the 1980s and perhaps indefinitely. The primary motivation for developing time sharing, however, is no longer as compelling as it once was. When time sharing was developed, economic considerations dictated that the only way that many people who had a need for a significant amount of interactive computing power could get it was by sharing very costly resources with many other users.

In the future we will see the rapidly increasing availability and use of "personal" computers that are as powerful as the most powerful machines of a decade ago, but which will sell for a small fraction of the cost of those machines. They will be usable in a stand-alone mode, but they will also communicate with other computer resources via networks and thus will provide users with facilities that are beyond the capabilities of personal machines. A common use of personal computers will be to support complex displays and word-processing applications locally, while depending on remotely located and shared facilities for long-term storage of programs, large number-crunching applications, and so on.

Taken together, the proliferation of information systems and their predominantly interactive character make it imperative that the human factors aspects of the design and operation of these systems be energetically addressed. Many past systems have failed for lack of good human engineering. Such failures must be avoided in future systems if the full potential of this rapidly evolving technology is to be realized.

1.6 Some Anticipated Changes in DoD Computer Systems

There is a growing awareness within DoD, and within the government more generally, of the fact that information technology is advancing at a sufficiently rapid rate that many current systems and operating procedures are obsolete or are fast becoming so. Evidence of this increasing awareness and recognition of the need to institute changes that will exploit this technology are found in a variety of recently issued reports. In what follows, we briefly review selected aspects of several of these reports that relate most directly to the subject of user-computer interaction.

1.6.1 The WWMCCS Information System and Its Planned Modernization

Computers have been widely used in military contexts since they were developed, and the dependence of military functions and operations on computing resources has increased dramatically over the years. A problem that was recognized twenty years ago was that of assuring the compatibility of the many systems that were being implemented by the different military forces and commands. A DoD directive issued in the early 1960s established the concept of a World-Wide Military Command and Control System (WWMCCS) and emphasized the importance of linking the various existing and anticipated military command and control systems to a national system that would support the National Command Authorities (NCA), which at that time consisted of the President, the Secretary of Defense, and the Joint Chiefs of Staff. While each command was authorized to develop systems to meet its own particular requirements, the intent was that these systems were to interface in such a way as to meet also the needs of the NCA. The approach did not work well; in particular, the computing facilities that had been developed to meet the needs of specific commands could not be made to work well in combination.

In the early 1970s, DoD adopted a new policy that centralized the planning of computer procurement and established certain standards and constraints regarding both equipment and procedures. The primary function of the WWMCCS was redefined to be that of supporting the NCA. The National Military Command System (NMCS) was designated to be the focal point of WWMCCS, and command and control systems of all DoD components were to be configured and operated so as to support the NMCS as well as their own specific missions.

To facilitate standardization and interoperability, two properties to which high priority was attached, the decision was made to build the WWMCCS around a single off-the-shelf model of a computer for which research and development were essentially complete. The computer that was selected was the Honeywell H-6000. As of early 1981, 83 H-6000 CPU's had been procured for use in the WWMCCS. These machines are deployed at 26 operational sites. Twenty of these sites are connected by means of the WWMCCS Intercomputer Network (WIN). As of 1981, the total number of work stations in the WWMCCS is in excess of 1,500 (1087 located at WWMCCS sites and 462 located remotely).

WWMCCS software falls into two broad categories: system software, which tends to be standard across sites, and applications software, which may be unique to one or a few sites. It has been estimated that there are currently about 8 million lines of system software code and about 18 million lines of applications software code. Much of this software is considered to be in need of redesign and modernization.

Recently (January, 1981), the Assistant Secretary of Defense (Communications, Command, Control and Intelligence) released a report describing a plan for modernizing the automated data processing capabilities of the WWMCCS Information System (WIS). The report was prepared for the U.S. House of Representatives Committee on Armed Services, in response to House Report No. 96-916. In what follows, the report will be referred to simply as the ASD WIS report.

The ASD WIS report contains much that is relevant to the problem of defining militarily significant research problems in the area of user-computer interaction. Mention of two specific problems that it identifies -- inadequate on-line software development and data management tools, and user-computer interface deficiencies -- suffices to make the point.

The ASD WIS report summarizes the need to modernize WIS as follows:

> a major modernization and enhancement of the current WWMCCS ADP and the entire WIS, including the basic information reporting system and its procedures, will be required over the next ten years to meet national priorities for situation assessment, crisis operations, and rapid deployment and support of military forces worldwide. Modernization of the ADP hardware alone will not be sufficient to provide the capabilities required for the wide range of WWMCCS functions. Redesign and modernization of the major applications software which supports a broad range of functions and users are essential. (p.14)

DoD's interest in modernizing the WWMCCS Information System stems basically from two facts: (1) WWMCCS ADP has become a vital resource to its users, and (2) there is a general recognition that the system is not as effective as it could or should be, in part because the hardware and software on which it is built are now out of date and lagging the state-of-the-art of information technology.

According to the ASD WIS report, most major U.S. military installations in the U.S., Korea, and Europe are connected to the WIS. Typical purposes for which the system is used include the following:

- maintenance of status and location of forces and resources,
- planning for force mobilization and deployments,
- preparation of the SIOP (Single Integrated Operational Plan),
- calculations for SPACETRACK,
- scheduling of MAC cargo and passenger reservations,
- estimating and monitoring Navy fleet fuel consumption,
- assistance in preparation and processing of AUTODIN messages, and
- assistance in preparation of Air Force tactical "frag" orders.

It is significant that the WIS Modernization Plan rejects the possibility of directly replacing obsolete machines with more modern equipment in a one-for-one fashion so as to preserve the overall system architecture. Instead, the decision has been made to develop a <u>distributed computing system</u> in which small computers dedicated to individual functions are linked by means of local networks. The anticipated advantages of such an architecture include flexibility and modularity. A distributed computing architecture is also expected to increase survivability of a system, inasmuch as the loss of one or a few components does not necessarily bring down the entire system.

Another key concept relating to the modernization plan is that of <u>families of functions</u>. The ASD WIS reoort identifies five such families, the first four of which are operational families, and the fifth of which is a general purpose family of command center functions. The four operational families are as follows: (1) resource and unit monitoring, (2) conventional planning and execution, (3) nuclear planning and execution, and (4) tactical warning and space defense. The general-purpose family of command center functions includes: (1) local network, (2) data base management systems for data base storage, retrieval, and manipulation, (3) security controls, (4) user support functions, (5) message handling, and (6) graphics support. The command center family would be required at every WWMCCS site whereas the operational families would not be. (The resource and unit monitoring family would be required at most sites). The planned capabilities of the four operational families reproduced from the ASD WIS report are attached as Appendix A.

The plan assumes that both hardware and software will be standardized within a functional family but not necessarily across families, although it is recognized that standardization is desirable, in general, for logistic training and maintenance purposes.

The ASD WIS report distinguishes two categories of WIS modernization activities: (1) those that relate to the functional families and are common to a number of sites, and (2) those that relate to command-unique requirements. Command-unique hardware and software would not be subject to the same standardization constraints as the functional families, but could be designed as appropriate to the needs of the specific command. Responsibility for the development of the hardware and software components to support the former activities will rest with the joint program manager. Command-unique activities will be the responsibilities of the individual commands.

The schedule for modernizing WIS identifies four phases, the first three of which are to start concurrently at the beginning of FY 1982. The first phase will be devoted to upgrading the existing system by correcting some of its major deficiencies so as to get some short-term improvement in performance. Phase II will focus on enhancing support to command center personnel. The emphasis in this case is particularly salient to human factors research, inasmuch as the intent is to make the system more available and useful to the user. One of the ways in which the usefulness of the system is intended to be enhanced is through the provision of user-oriented automated message handling. This includes "supporting functions such as the composition, coordination, and transmission of messages developed by users and the automated receipt, distribution and accounting of messages received by the users. Other functions such as maintenance of historical message files, on-line preparation of private user-oriented files, and gathering of statistics will be included" (p. 41).

Inasmuch as the primary mode of use of WIS is intended to be an interactive one, the design of the user work station and of the software that supports the user-computer dialogue are also critical issues. The intent, according to the plan, is to develop a family of work stations, each member of which would provide a different level of capability for the user. The plan appears to be to implement work stations that initially provide the user with access to the automated message handling function, to the network interface and to selected applications supported by the H-6000s of the current system, and to expand the capability of the work station so as to provide access to additional WIS functions over a period of time, eventually providing access to all WIS functions through at least some of the individual stations.

Phase III of the WIS modernization effort will be devoted to the development of the software necessary to support each of the four operational families of functions. Phase IV, which is scheduled to begin late in FY83, will focus on development of the hardware and software required to support command-unique requirements. The initial plan carries the effort through 1990 and the total cost is estimated at between $1 and $1.3 billion.

A particularly noteworthy aspect of the ASD WIS report is its recognition that the conventional model of the system design process, in which one moves sequentially from a needs assessment to a functional specification to a hardware design to an implementation, is unworkable for complex computer-based systems. A major reason for its unworkability is the fact that the introduction of complex and powerful tools often changes the character of the functions that are performed and the work that gets done. The report puts it this way:

> A further consideration in the requirements determination process is the fact that the introduction of ADP has already had (and is expected to continue to have) a major impact on the way in which command and control is organized, broken into functions, and

> implemented. Thus, requirements for command and control systems involving ADP do not fall neatly into the traditional "top down" process. A serial approach of first stating requirements and then undertaking architectural aspects cannot be rigidly applied here (or in other systems with a high dependence on organizational and human aspects). Rather, a parallel requirements-architecture and feed-back approach is indicated, with interaction and feedback at each level of the hierarchies. (p. 29)

One of the implications of this view for human factors research is the need for its involvement throughout the system design and implementation process. If the introduction of increasingly sophisticated computer-based tools into military operations will change the character of those operations, and in particular the nature of the demands that are placed on the users of those tools, it is essential that the kinds of studies be performed that will anticipate those changes and assure the user's ability to adapt to them.

1.6.2 CENTACS and the Military Computer Family

In keeping with its policy of standardizing on computer equipment and assuring that all equipment procured meets certain Army requirements, the Army established in 1974 the Center for Tactical Computer Systems (CENTACS) to provide a focal point for the development of all of its tactical computer-based systems. According to the Standard Computer Resource Interface and Management Plan (referred to as the SCRIMP report) issued by CENTACS (1980), CENTACS' activities focus on five major areas:

- The Military Computer Family (MCF);
- The DoD programming language designate, ADA;
- The Intelligent Terminal Family (ITF);
- A Teleprocessing Design Center (TDC); and
- A Systems Management Engineering (SME) function.

The Military Computer Family will include a super-minicomputer (AN/UYK-41), a microcomputer (AN/UYK-49), and a single module (card) computer, all of which are to meet Army-developed standards. The intent is that all battlefield automated systems will make use of members of the MCF after they become available in 1987. The General Specification (CR-CS-0037-001, June 1980) calls for human engineering of equipment designs in accordance with MIL-STD-454 requirement 62. It mandates that "maximum effort shall be directed toward reduction of human operational error, particularly in the design of formats/information displayed to the operators and maintainers" (p. 12). While the computers of the MCF will vary in size and capability, all will make use of the ADA programming language and will feature a common instruction set architecture. This instruction set architecture, which is called Nebula (MIL-STD-1862, 28 May 1980), is a 32-bit architecture designed for efficient compilation and execution of ADA programs.

ADA is the product of a five-year effort to design a high-level programming language to be the standard language for military use. Design specifications went through a series of revisions, the code names for which were STRAWMAN, WOODENMAN, TINMAN, IRONMAN, STEELMAN, and STONEMAN. ADA is scheduled to be introduced into Army operations in 1983.

The Intelligent Terminal Family refers to terminals, displays, and other peripheral devices that will be used to provide access to computing resources. One terminal that has received much attention to date is the so-called Digital Message Mini Terminal, a general purpose hand-held message entry device. What the Army wants in this case is a terminal of less than 60 cubic inches, weighing less than two pounds, that will provide interactive capability for message generation, editing, storage, retrieval, and display. Prototype models of such a terminal are currently being developed under contract to the Army by the Magnavox Corp.

The SCRIMP report identifies the lack of a standard product line of terminals as being responsible for "a growing proliferation which is impeding the attainment of interoperability, continuity of operations, security, reliability, availability and maintainability" (p. 11). It also identifies the lack of a map background and the inability to interact with a data base as inadequacies of the existing digital plotter map and electronic tactical display used in the Tactical Fire Direction System (TACFIRE).

The Teleprocessing Design Center is a computer laboratory that contains, among other things, a Microprogrammable Multi-Processor System, which is used to emulate other systems (such as the TACFIRE and the TTC-39 Programming Support System) for purposes of debugging and analysis.

The purpose of the Systems Management Engineering function is to provide system engineering support and coordination so as to ensure the compatibility of the various resources that are developed for the MCF.

According to the SCRIMP report, both military doctrine and emerging technology point up the potential value of designing and implementing future military information systems as distributed systems. The benefits of distributed systems, the report suggests, are at least three: improved system performance, survivability, and flexibility.

This emphasis on distributed systems is consistent with that found in the plan for modernizing the WWMCCS Information System. It is also a reflection of a major current trend in information technology more generally, and it has definite implications for the kinds of human factors problems that will be encountered in the systems of the foreseeable future. We will return to this theme several times in this report.

1.6.3 DARCOM's Blueprint for Information Processing in the 1980s

The Army Material Defense and Readiness Command (DARCOM) is looking to the introduction of new information technology in the 1980s to improve the "metabolism rate" for its operations and management. In the words of a recent report from this agency, entitled "Blueprint for DARCOM Information Processing in the 1980s," "the infusion of technology should shorten completion times for actions of all kinds including supply transactions, development cycles, research projects, procurements, management and budgeting processes, and administrative actions" (DARCOM, 1980, p. 3).

The Blueprint identifies eight different types of needs for information processing and communication:

- General needs: the need to provide continuity of current operations.
- Personal support: e.g., support needed to provide personnel with individual access to various systems for informal electronic communication, document preparation, information management and other office functions.
- Transaction handling: the need for interactive systems to facilitate origination, submission, and follow through on transactions.
- Narrative text handling: tools for text prepartion and editing, document assembly and distribution, storage and retrieval of textual information.

- Technical data handling: tools for dealing with large amounts of numeric tabular data.
- Graphic data handling: the need for tools to facilitate preparation and use of graphic materials such as technical drawings, and to integrate graphic material with narrative text and documents.
- Filing and retrieval: the need for central mass storage to retain very large amounts of information, and access to those repositories via electronic data networks.
- Information systems interconnection: to provide access to the same information from multiple sources and to assure information mobility.

The DARCOM Blueprint points out the importance of having text editors, message systems, and information retrieval systems functioning in an integrated way. One wants to be able, while preparing a document, to retrieve needed information from various data banks using the same terminal for searching for the information and for editing it into the manuscript in preparation. Similarly, the same terminal and, preferably, the same computing environment should provide the user with the capability of sending the document draft or parts thereof to colleagues for review or for purposes of collaboration in its preparation.

The report also identifies the following six major thrusts that are being undertaken to improve the support that computers are providing for DARCOM's mission:

- The architecture thrust: to move particular software functions from existing mainframe computers (IBM and CDC machines) onto minicomputers that will be connected to the existing mainframes; when the workload of the mainframes has been reduced by approximately 50%, a decision is to be made regarding whether to continue transferring workload to the minicomputers or to acquire new large mainframes to service some portion of the load.

- The distributed processing thrust: to dedicate minicomputers with their own local data bases to the support of selected mission areas.

- The narrative processing thrust: to provide text editing tools, message systems, and other office automation tools on personal computers.

- The digital technical data system thrust: to provide new approaches to the handling of technical data packages including the digital representations of technical drawings.

- The security thrust: to increase the physical and electronic security of computing facilities, commensurate with the increasing dependence of DARCOM on their use.

- The networks thrust: to service the increasing requirement for exchanging large volumes of data at various speeds among the many computers in the command.

These thrusts have been identified for the purpose of permitting an evolutionary improvement in the use of computing resources within DARCOM without causing an upheaval in the command. The smoothness with which the intended improvements will be realized will depend in no small degree on the attention that is given to user and interface issues in the implementation of these thrusts.

1.6.4 Report of Advisory Committee on Information Structure and Functions for the EOP

In May of 1979, the Director of the Office of Administration of the Executive Office of the President (EOP) established an Advisory Committee on Information Network Structure and Functions. (Although the EOP is not part of DoD, its plans for implementation and use of computing resources are clearly germane to the purpose of this report.) The Committee's purpose was "to outline a structural and functional plan for the EOP information network, striving for immediate implementation and minimum network life of 10 years" (Dertouzos, 1980, p. 6). The EOP Information Network was defined "as consisting of all non-mission oriented information processing and data communications services along with supporting hardware and software systems that are centrally administered and common to the EOP user community" (ibid., p. 12). The intent of the developers of the network is that it provide access to any computer resource by any authorized user on a 24-hour-a-day, seven-day-a-week basis. Among the administrative units that would be expected to make use of the network are the Executive Office of the President (EOP), the Office of the Vice President, the Office of Management and Budget, and the National Security Council. In addition to providing reliable and convenient service, the network is intended to be able to grow gracefully with the technology to insure the privacy and authentication of data and to be in the technological forefront of fiscally prudent systems.

The services that the Advisory Committee recommended that the Network initially provide include the following:

> (1) a top-level command language for accessing all network core services; (2) means for providing communication between computers, terminals and other network nodes; (3) means for user login and logout; (4) a text editor and formatter; (5) means for the preparation, transmission and processing of messages among EOP users; (6) means for transferring large data files from one network to another; (7) means for performing speed, code and protocol conversion. (p. 3)

Services that are expected to be added to those listed above include:

> (1) means for mixed media communications; (2) extended office-related functions; (3) means for transparent use of the network, e.g., by users requiring very high-speed links; (4) an on-line tutoring service for assisting novice users; (5) means for text format translation, e.g., for transferring data among different word processors; and (6) hardware/software privacy and authentication mechanisms. (p. 4)

It is clear that the development of these services in such a way as to ensure their usefulness and usability will involve dealing with a number of human factors problems, e.g., determination of the characteristics of the top-level command language, the text editor and formatter, the message system, the file transfer protocols, and the on-line tutorial help features.

All of these things relate to the user-computer interface; the ease and effectiveness with which users will be able to interact with the system will depend heavily on the specifics of their design.

1.6.5 Report of the IDA on C^3I Data Communication Networks

In December, 1977, the Institute for Defense Analysis (IDA) was asked by the Office of the Director, Information Systems, Assistant Secretary of Defense (Communications, Command, Control and Intelligence) to analyze the computer-data communications networks currently in development, especially with a view to their implications for the intelligence community. The conclusions and recommendations in the final report from the study (Bartee, Buneman, Gardner, and Marcus, 1979) relate to three major topics: (1) communications systems protocols, (2) query languages for using data base systems, and (3) the advisability of a local Washington area data communications network.

Of particular relevance to this report are some of the conclusions and recommendations relating to query languages for providing network access to data base systems. One such conclusion the investigators drew was that there currently is no language that is completely satisfactory for all C^3I users. The need for a common language, or at least for some standardization among languages, derives in part from the fact that there are currently at least 100 different data bases that are used for intelligence purposes, which are maintained by upwards of 20 different systems. Given that each system has its own query language and that most query languages are not trivially easy to learn to use, accessibility of these data bases to the average potential user is seriously limited. The authors point out the need not only for standard ways of making specific queries of the systems, but also for browsing capabilities.

It is of some interest that the authors of the IDA report question the advisability of putting great emphasis at the present time on the development of natural language interfaces to intelligence systems. They argue not that this possibility should be ignored, but rather that there are higher-priority problems that require resolution. Following a review of several query languages based on different data base structures (e.g., relational, network, hierarchical), the authors conclude that what is needed is the development of "intermediate" query languages into which several other query languages could be readily translated. One language that has been proposed to serve the function of an intermediate language is ADAPT (Glaseman and Epstein, 1978). While the authors of the IDA report agree that ADAPT should be developed further with this intermediate language role in mind, they recommend the exploration of other possibilities as well. They also recommend that attention be given to the design of a friendly interface for whatever intermediate language is developed. In addition, they note the desirability of developing techniques to permit the querying of a large data dictionary that encompasses data bases at different geographic locations. Such querying could constitute a top-level browsing through the entire collection of data bases, looking for pointers that could be worth pursuing.

1.6.6 Summary Comment

The reports mentioned above were produced by several different authors and agencies. Each presents a somewhat different perspective on the use of computers by DoD in the near-term future, but they share certain common themes. From these reports, one gets a picture of anticipated radical changes in the way computers and computer systems will be structured, deployed, and used. An obvious trend is in the direction of distributed operating systems. Rather than being concentrated in a few large, remotely-accessed mainframes, computing power will be widely distributed among a very large number of small but very powerful machines. These machines will be interconnected via data communications networks, and networks themselves will be interconnected via internet gateways. Large machines will still be used, but they will not be the source of much of the computing power that is required at operational sites; they will be the repositories for very large data bases and the work stations for problems that require excessive amounts of computing power.

The cost, standardization and interoperability of software will be of increasing concern. The cost of software development is already considerably higher than the cost of hardware, standardization is minimal, and interoperability is practically non-existent.

The tools that will be available to end users will be numerous and sophisticated. Among the kinds of tools that appear to be of special interest are those that relate to message preparation and processing, and to information management more generally.

The prevailing mode of computer use will be interactive, and the user community will include people at all levels within the command structure, most of whom have had little, if any, training in computer technology. Consequently, there is wide recognition of the need for "friendly" interfaces, user aids, and effective safeguards against human error. The development of versatile terminals that provide adequate input-output bandwidth and require relatively "natural" inputs from the user is also a perceived need, as is the development of query languages and interaction protocols that will facilitate effective communication between user and machine.

The increasing dependence of the military on computer-based systems, and on information technology more generally, is obvious. That the systems currently in operation are rapidly becoming obsolete appears to be widely recognized, as the emerging plans for modernization attest. Any modernization effort that is to have a chance for success must pay careful attention to issues of usability. Given the newness and complexity of the architectures, tools, and capabilities that are being developed and our lack of experience in using them, many of these issues are opportunities, if not mandates, for human factors research.

2. MILITARY APPLICATIONS OF INTERACTIVE COMPUTER SYSTEMS

To provide a background for the discussion of specific issues in the following sections, we discuss briefly here a few activities and functions that are relevant to Army operations as they are currently performed or as they are likely to be performed in the near future. It is not our intent to attempt anything like a comprehensive survey of military operations, but rather to touch upon a few activities that we believe are illustrative of the variety of ways in which computers are, or soon may be, used in military contexts.

2.1 Communication

The centrality and criticality of communication for military operations is obvious. Without effective communication between and within operational units, the military could not function. This is true in peacetime as well as in wartime; effective communication is important to the objective of avoiding conflict and crisis as well as to that of resolving conflict or managing crisis. The increasing complexity of military systems involves a paradox with respect to this problem: on the one hand, the communication technology is becoming ever more powerful and versatile; on the other hand, the demands for rapid situation assessment and decision making and for the coordination of geographically and organizationally separated activities are ever more severe.

An apparent trend in military communications is the increasing dependence on computers. Indeed, the extent of this dependence has become so great that the distinction between computer systems and communication systems is increasingly difficult to maintain (Nickerson, 1980). It is worth noting in this regard that the ASD WIS report also identifies communication between personnel at different command centers via on-line teleconferencing and message exchanging as a significant use of the WWMCCS Intercomputer Network.

A concept that could have profound implications for the need for communications is that of the distributed command post. The proposed concept calls for dividing the command post into functional cells, each of which is independently mobile and has self-contained communication capabilities. Cells would be dispersed at intervals such that no single conventional weapon strike could take out more than one of them. The principal advantages of the concept are that a distributed command post would be difficult to find, hard to disable, amenable to incremental replacement, and easy to move.

While not yet approved as official Army doctrine, this concept is compelling, and it appears that some mechanism for decentralizing command post operations will become essential as electronic detection and targeting systems and other battlefield hardware become more sophisticated. A centralized divisional command post with 200 to 300 personnel is simply too easy to identify and too significant a target to remain viable as technology improves.

In what follows, we discuss briefly several aspects of communication operations, illustrating their importance at all levels of command, from foot patrols through battlefield command posts to the National Command Center.

2.1.1 Message Processing

As was noted previously, rapid, reliable communication is essential to any group that must act in a coordinated fashion while dispersed geographically. Effective, continuous communication is especially critical during rapidly changing situations. One frequently hears reports, however, of communication failures in military situations that would seem to have been preventable. A common complaint is that message centers often find it impossible to deal with the amount of message traffic that some tactical situations produce. When the center is loaded beyond its capacity, important messages may be delayed, misdirected, or lost completely. The development and use of message processing techniques that would assure the timely, accurate routing of messages and minimize the probability of a system breakdown due to overload is a continuing military problem of considerable significance.

Consideration of this problem suggests that there are several underlying generic issues. What rules and procedures should govern the sending of messages? Should it be allowable to address messages to any possible recipient, or should the message flow be constrained to follow hierarchical organizational lines? Should special communication procedures be followed in a crisis? One of the characteristics of a computer-based message (electronic mail) system is that it provides all users with access to all other users. Is this a desirable thing in a tactical communication system? If it is not, what kind of an access control discipline is required? To what extent should this discipline be changeable by command or adaptable to situational demands? In a computer-based message system, there is the possibility that knowledged-based or other advanced techniques may be employed to help manage the flow of messages so as to minimize the impact of crisis situations.

Another set of issues relates to the kind of feedback that a user of a tactical communication system should receive: feedback regarding the status of the system (is it alive or dead? is it attending to that particular user?), feedback regarding whether a message has been sent (and received), and feedback regarding whether the user's intentions have been realized (that is, whether the prescribed actions have been taken).

Many of the messages that are sent over a tactical communication system are of interest or use only for a short period of time. Such messages should (presumably) be purged from the system when they are no longer useful. On the other hand, some messages must be retained for long periods of time, if not indefinitely. What procedures and policies might be developed to assure that a system retains the information that should be retained and automatically purges what should be purged, and purges it as soon as possible so it does not represent an unnecessary drain on system resources?

The more sophisticated computer-based message systems now provide users with many more functions than simply those of sending and receiving messages. In particular, they are beginning to provide the functionality of local information management systems. Like most other software systems, computer-based message systems have been developed in accordance with the intuitions of their designers. Their development has not been guided, for the most part, by empirically validated principles about how communication and information management functions should work.

2.1.2 Content-dependent Message Routing

A primary purpose of a communications system is to transmit messages from one point to another. The destination of a transmission is typically specified by the sender. For many purposes this addressing policy works very well because the sender knows to whom the message <u>should</u> be sent. For some applications, however, sender-controlled addressing does not work so well, and a preferred policy is to have receivers designate the kinds of messages they wish to receive. Intelligence information-distribution systems sometimes work on this basis. An intelligence specialist may specify, for example, that he wants to receive all incoming messages or reports that pertain to his area of specialty (e.g., all reports of activities of a certain type or occurring in a certain geographical region).

In a computer-based message switching system, it should be possible to do some amount of message routing automatically on the basis of message content in accordance with the interest profiles of receivers. How much of this can be done, and how effectively, depends on several factors, especially on the ability of the system to analyze messages according to content. Manual classification of messages with respect to content by senders is, of course, a way of avoiding the difficult problems of automatic content analysis. However, this approach requires sophisticated senders who are knowledgeable with respect to the interest categories of the potential recipients. Automatic classification with respect to content, if it could be done with acceptable accuracy, would permit the efficient use of operators whose training and knowledge are at considerably lower levels than those required for manual content analysis.

2.1.3 Teleconferencing

One of the primary questions being debated in connection with the concept of the distributed command post is the question of how the lack of face-to-face contact among command post personnel might degrade staff coordination. This question has many facets. The existing structure is the product of at least 200 years of evolution, and it will not be discarded lightly. On the other hand, it is clear that demands of modern warfare are severely straining the capabilities of the command structure to assimilate and assess information, to evaluate alternatives, to formulate plans, to communicate them to all necessary units, and to coordinate their execution in timely fashion. The recognition is growing that victory on future battlefields may well belong to the side with the shorter overall update time for this cycle.

Regardless of the outcome of the debate over the distributed command post concept, the requirement for mobile forces around the world has already made multiple-person communication via remote telecommunication links an essential component of future Army operations.

Over the last 20 years, there have been episodes of intense research on the design of such teleconferencing capabilities, but no sustained effort leading to systematic principles of design. In 1963, the Institute for Defense Analysis undertook several interrelated studies concerned with evaluating the feasibility of high-level international conferences via telephone or teletype (Bavelas et al, 1963). In the early 1970s, Klemmer, at Bell Telephone Laboratories, edited a special issue of Human Factors (Klemmer, 1973), specifically concerned with interpersonal communications using advanced technology. Many of the groups identified in that issue continue to be active researchers in the field. Chapanis, et al (1972, 1977) have reported a series of studies evaluating alternative communication modes for two-person problem solving and; similar studies have been done on alternative schemes for permitting larger groups to interact through telecommunications systems (Forgie, Feehrer, & Weene, 1979).

Many of the same issues that have been the subject matter for this work remain, such as:

- identifying the maximum size of a teleconference that will maintain operational effectiveness,
- evaluating the impact of transmission delays,

- evaluating the contribution made by voice, video, and hard-copy transmissions among conference sites,
- central vs. distributed control of the conference, and
- defining appropriate training strategies to enhance user effectiveness.

Other issues have evolved as a result of the requirement, particularly in military communication, to maintain cryptographic security. It is now necessary to research difficult questions having to do with reduced bandwidth communications, such as speech intelligibility and acceptability, speaker recognizability, message integrity, and turn taking.

These questions do not have absolute answers, and they must be addressed in relation to a variety of contexts. It is also to be expected that there will be interaction among a number of independent variables. For example, the size of an effective conference surely depends on whether substantial transmission delays are involved. It also depends on the number of modes of communication that are available.

The proliferation of computers and computer networks has added a new dimension to interactive media. It is now possible to envision multimedia workstations, incorporating voice capabilities and CRTs, that provide access to people and data bases distributed around the network. From a technical standpoint, it would be feasible to design a work environment in which face-to-face contact would never be required. The need for such contact would depend on the extent to which effective interpersonal functioning required it. To our knowledge, aside from space-to-ground communications over extended periods, such circumstances have never been researched to determine both the work-environment design issues and the social-psychological issues that will impact on remotely-coupled team performance.

There are at least two types of effects that have human factors implications for teleconferencing. The first type involves the effects of bandwidth limitations on interpersonal communications. Assuming that all equipment is working according to specifications, what bandwidth is really necessary? How important is it to be able to transmit detailed pictorial information? How important is it that color information be included, and that the color be accurate? The second type involves the effects of degradation in the communications process. What happens when participants in a conference drop out (or in) unexpectedly? What are the effects of uncertainties in whether one's latest transmission has been heard or not?

There are other issues that go well beyond the mechanics of the teleconferencing process. Any significant change in command post organization will necessitate the development of new protocols, procedures, and techniques, and some time will be required before these are optimized for a given set of communications capabilities. The command structures themselves may need to be modified in the light of the new procedures, and additional staff organization may be needed. Final judgments must be deferred until the intertwined relationships reach a reasonable equilibrium.

Other fundamental questions that should be addressed include the following: How important is speaker recognition in a teleconferencing situation? How important is it that the speaker be recognizable on the basis of voice cues? What, if any, alternative identification or authentication methods are acceptable?

2.1.4 Mobile Digital Communications (Packet Radio)

We noted in the previous section that one of the anticipated major developments in the use of computer technology in the military is an increasing emphasis on distributed operating systems and distributed command systems. This implies, among other things, computer systems linked via data communication networks.

One of the advantages of a distributed-system architecture is increased accessibility to system resources from many locations. Ultimately, a goal would be to make these resources accessible from anywhere -- even by users who are on the move. A problem that must be solved to provide this kind of access is that of giving the individual user a wireless connection to the nearest available node on the network. One potential solution to this problem is packet radio.

Packet radio and its potential use in military contexts are currently being explored. The development and use of this technology raise a number of human factors issues and problems. Moreover, these issues and problems will become even more relevant as the technology progresses to the point of making radio transceivers sufficiently portable that they do not require a vehicle for mobility but can be carried by individual personnel. Specific issues include:

- The design of a versatile hand-held terminal. The problem is to design a terminal that provides sufficient bandwidth (both input and output) and does not require an inordinate amount of training (as does, for example, a stenotype machine).
- The design of languages or codes that can be used for communication purposes in a packet radio network.
- The possibility of speech in the packet radio context.
- Issues of operator training and equipment maintenance.

2.1.5 Foot Patrol Communication

Military communication problems range from those associated with disseminating information from the National Command Center to military units deployed around the world to those of the infantrymen in a foot patrol who must keep in contact with each other without divulging their presence to enemy forces. While there is a natural tendency to focus on the higher level problem because of the far-reaching implications of communication at this level, the problem of the foot soldier is equally real and, at least from his point of view, of considerable importance.

Brown (1967) has argued that while communication is one of the army's most perplexing and pressing problems in general, it is at the level of the small unit and in the context of the type of patrolling actions that were conducted in Vietnam that the problem becomes most acute. Both auditory and visual signalling schemes are used in such situations; however, both have limitations. Auditory communication is sometimes precluded by battlefield noise or by security considerations. Visual hand-signalling is useful only under conditions of unobstructed line-of-sight between sender and receiver, and during daylight. Tactile communication -- via either direct touch or remotely transmitted electrical signals to the skin -- offers some advantages, relative to either audition or vision. (Another context in which tactile communication can be used to advantage is that of underwater settings. Visual signalling schemes have also been developed for this purpose, but, again, they are usable only under adequate lighting conditions.) A human factors problem associated with tactile communication is that of designing an adequate coding system from the point of view of both sender and receiver.

2.1.6 Reporting of Location Information

A type of information that is often important to transmit and to transmit accurately in a tactical situation is location information. The most common way of reporting such information is verbally, by giving coordinates. It should become possible with the use of computer terminals in tactical situations to designate location by pointing on an electronically displayed map. Whether designating location in this way would be less error prone than the verbal reporting of coordinates is an empirical question.

2.1.7 The Problem of Communication Gaps or Failures

It is apparent that communication failures of various sorts can be catastrophic in crisis situations. It would seem to be imperative, therefore, that communication systems and procedures be designed so as to function well, if not optimally, during crises. Ironically, it is during just such times, according to some writers, that communications are most likely to break down. Greene (1973), for example, points out that established channels and formal information-distribution procedures are often worthless when unanticipated crises arise. "The logic of distribution often breaks down during crises and commanders spend a good deal of their time trying to find out where their message should be sent" (p. 111). Again, "information systems procedures and displays which are based on some sort of format which is amenable to automatic data processing are almost useless during these [crisis] times" (p. 112).

Greene stresses the nonroutine and unpredictable nature of the messages that must be sent during crises and the need for flexibility, both with respect to message routing and message format. Communication gaps that may result from the inability to cross organizational boundaries in unanticipated ways are noted as a special problem. Greene raises the following question, which has both technical and political implications: "Can an information system be devised which would facilitate a communications flow between offices in different chains-of-command without threatening the authority of those who are 'higher' in the organizational structure?" (p. 112). The emerging technology of computer-mediated communications systems is seen as a promising development that may facilitate the exchange of information both within and across organizations; it may introduce other problems, however, associated with security and information privacy.

One approach to the study of communication failures has been through mathematical modeling of communications systems. This approach is illustrated by the analysis of communication gaps by Johnson and Mirchandani (1976). The object of study in this case was a communication network composed of multiply interconnected nodes. The point was to derive, for any two nodes in the network, the probability of connectivity, the transmission delay time, and the degree of distortion in transmitted information, given the values of these variables for all two-node links that were contained within any path between the nodes of interest. This approach produces measures that represent overall performance of a system in an aggregate way, assuming accurate estimates of the values of the output variables, and can be very useful in determining how the nodes of a network should be interconnected to assure a desired level of performance.

2.1.8 Communication Security

One of the key problems in military communications systems is that of security. While maximizing the ability to maintain communication with friendly forces, one wants at the same time to minimize the possibility of interception by, and of revealing one's location to, hostile forces. As we have already noted, a particularly difficult communication problem is that of maintaining contact among members of a foot patrol in such a way as to permit some dispersion and at the same time preclude detection.

2.1.9 Communication System Criteria

The problem of establishing criteria for judging the adequacy of a tactical communication system is a difficult one. There are certain obvious desired properties, such as reliability, adequate bandwidth, security, survivability, and mobility. It is not apparent, however, how these factors should be weighed in arriving at a single figure of merit. Moreover, there are other factors that relate more directly to useability that are also critical. How well does a system meet the needs of its users? Does it get the *right information (and not a lot of* the wrong information) to the right people when it is needed? Does it permit its users to interact with it in a reasonable and relatively natural way?

2.2 Situation Assessment

One of the fundamental problems of military operations is that of situation assessment. This activity requires the gathering of information from many sources having different degrees of reliability and completeness, evaluating that information, integrating it into a coherent whole, and inferring from it the details of the current situation. In a computer-based system, the problem includes that of assuring the currency of the data in the data base (or assessing and coping with its non-currency or unreliability), finding a useful way of organizing the data within the computer, finding an effective way of representing it to a user, and giving the user some control over this representation so that he can selectively interrogate the data base and view different aspects of the situation. These concerns involve a number of fundamental human factors problems, including data base organization and the problem of imperfect information, command languages, display presentation issues, and user models. (These topics are discussed in Sections 3 through 5.)

2.3 Office Automation and Information Management

When one thinks about military applications of computers, office automation functions are not the first things that come to mind. However, the ASD WIS report (1981) notes the growing use of WWMCCS ADP for a variety of command center support functions that are similar to what the commercial world describes as office automation. This includes the managing of records, logs, briefings, and messages. We believe that this type of function will continue to increase in importance both in military and non-military contexts. While many of these functions currently are discussed under the rubric of office automation, we prefer the term information management because we believe that to be more descriptive of the functions that are actually performed. The introduction of electronic tools to the office workplace does not automate the office in the sense of making human involvement either unnecessary or routine; indeed, the possibility that these tools will decrease the degree of automaticity of the jobs performed by human beings in the office rather than to increase it seems very real. They also have the potential of increasing the productivity and efficiency of the human beings' work very substantially.

2.4 Monitoring and Supervisory Control

Many of the day-to-day problems with which the military must deal in order to maintain a smoothly functioning operation may be characterized as problems of monitoring and supervisory control. This concept -- monitoring and supervisory control -- is a generic one that has been gaining increasing currency as a consequence of the changing nature of many jobs with the introduction of automated equipment and techniques in the workplace. One of the effects of increasing automation has been to create more and more jobs in which the operator is given very high levels of responsibility and very little physical work to do. The degree of responsibility and the amount of work vary from position to position, but the defining properties of such jobs are:

- The operator has overall responsibility for control of a system that, under normal conditions, requires only occasional fine tuning of system parameters in order to maintain near-optimal performance.

- The major function performed is to serve as a back-up in the case of a system component failure or malfunction.

- Important participation in system operation occurs infrequently and at unpredictable times.

- The time constraints associated with participation, when it occurs, can be very short, of the order of a few seconds or minutes.

- The values and costs associated with operator decisions can be very large.

- Good performance is associated with rapid assimilation of large quantities of information and the exercise of relatively complex inference processes.

These kinds of jobs occur in process control industries ranging from chemical plants to nuclear power plants. They are also involved in the control of large ships and urban rapid-transit systems, and they can be expected in medical patient-monitoring systems and perhaps in law-enforcement information and control systems. As computer aids become more widely used in military command and control systems, problems relating to supervisory control will become a matter of increasing concern in this area as well.

Supervisory control tasks are fraught with human factors concerns. There are few, if any, good principles for job design for operations of this kind, in part because it has proved extremely difficult to measure or estimate the workload involved. Without suitable attention to job design, the workload is highly variable, ranging from routine and boring when the system is operating smoothly to extremely demanding when it is not.

In the past, approaches to supervisory control have utilized large arrays of meters and gauges or large situation boards to display information. The strategy has been to display everything on the assumption that one never knows exactly what will be needed. Little attention has been paid to the need to assimilate the diversity of information from various sources into coherent patterns for making inferences simply and directly. Today, computers are being used more and more in the control of these operations, and the large display panels are being replaced with CRT displays with provisions for calling up the needed information on demand. However, these developments are pushing the state-of-the-art with respect to optimum ways of displaying and coding large collections of information for ease of interpretation. We also lack well-developed methods for identifying the conceptual models that operators evolve when using these systems, and thus for identifying suitable ways of presenting processed information to promote effective decision making.

Issues of training and proficiency maintenance also become important in this kind of operation because each critical event is in some sense unique and is drawn from an extremely large set of possibilities, most of which will never occur during the operating life of the system.

3. USER ISSUES

3.1 Person-Computer Function Allocation

As computers play increasingly large roles in today's systems, the question of exactly what functions should be assigned to people and what functions to the machine has become more and more significant and urgent. The issues are both methodological and substantive. At a methodological level, we need to have a collection of standard procedures that are useable by system designers with no specific training in human factors. These procedures should ensure that the assignment of functions will take account of human performance capabilities and limitations. At the substantive level, there is no consensus and no data on which to build a consensus concerning the fundamental design philosophy that should underlie the allocation. Should systems be designed so that users serve only as a back-ups? Should computers be allowed to choose and execute courses of action?

Taking the substantive issue first, one would obtain general agreement with the statement that people should be assigned the functions at which they excel and computers the functions at which they excel. However, this is a vacuous statement, not only because it may be difficult to reach agreement on what people are good at, but also because there remain some underlying philosophical issues that are equally important and that do not admit to this level of analysis.

A position that is sometimes taken is that one should examine the system requirements carefully and automate everything that it is technologically feasible to automate. The role for the human operator is to do leftover tasks, and, perhaps more importantly, to provide the back-up manual mode in case one or more of the automatic functions fails. A variant of this approach is to advocate automating any function whose urgency is such that that an operator cannot be expected to accomplish it fast enough.

A second position is that the human operator should remain in control of the process at all times. The computer may provide advice and recommend courses of action for the operator to follow, complete with step-by-step reminders of what to do. Advocates for this position argue that regardless of the process being controlled, there are always intangible judgmental factors that come into play that cannot be anticipated.

A third position is that the role of the computer should be limited to clerical and computational tasks and to tasks that make possible the generation of "intelligent" status displays. According to this position, the computer is neither monitoring the operator nor is the operator doing nothing more than monitoring the system. The computer provides integrative information and planning aids that assist the operator in understanding the system state or battlefield conditions. This integrative information may be knowledge-based in the sense that it is sensitive to the current operational context and takes advantage of an underlying physical or conceptual model of the system.

Each of these positions has been advocated by one or more individuals or organizations. Choices among these kinds of alternatives typically depend on political and social issues as well as technical considerations. It is unfortunate, however, when decisions are made on political grounds that have not been informed by evidence of the effectiveness of operator or crew performance under alternative allocation philosophies.

With respect to methodology, in many design projects the assignment of functions to people and to computers is made by default. Either traditionally human roles are assigned to people, or the decision is made to automate all possible functions and only those tasks for which automation is found to be infeasible are left to them.

Among the human factors community, formal task analyses are frequently completed to describe the work as it is done before computer assistance is introduced. An analysis of the characteristics of the user population is undertaken. Lists of characteristic advantages of human information processing and computer processing are consulted, and the final recommendations are the result of judgment informed by these analyses.

More systematic analytic and quantitative tools are needed to insure the consideration of human factors issues in design and to provide a formal way to allocate functional requirements among system elements. Pew, Woods, Stevens, and Weene, (1978) suggested a pseudo programming language specification of the task to be performed that might help to formalize the process beyond standard task analysis. Reisner (1977) showed how formal grammars made it possible to evaluate alternative designs and Pew, Sidner, and Vittal (1980) proposed the use of a knowledge representation language to represent the users' and the programmers' perspective in a common framework. Finally, Moran (1978) proposed a full-scale design language that would include specification at the task level, the semantic level, and the level of specific hardware interaction.

There is a serious need to develop and evaluate alternative approaches to function allocation that encourage careful consideration of user and system requirements and that are compatible with the way systems designers think about their work.

3.2 Types of Users

With the growth of interactive computing in the late 1960s, significant numbers of nonprogrammers began using computer systems. This development produced great pressure for improved user-computer interfaces, and many innovations resulted. Most of these innovations arose from within enclaves of programmers in universities and other centers of advanced technology and spread rapidly from one such community to another. As these innovations were integrated into systems designed for use by nonprogrammers, however, they sometimes failed to yield the anticipated effects. It seemed that many aspects of user-computer interfaces that were considered obviously desirable by experienced programmers were anything but obviously desirable to nonprogrammers. Moreover, it became apparent that nonprogrammers would not put up with the kinds of arcane inconsistencies that most programmers could tolerate.

Increasing attention was paid, therefore, to the human factors of user-computer interfaces. Gradually, it was recognized that there are several categories of users of computer systems, each of which has somewhat different characteristics and needs. These categories include what have been called "casual," "computer-naive," "intermittent," "nondedicated," "nonprofessional," and "inexperienced" users. It should be noted that these terms are by no means synonymous; on the contrary, they reveal the existence of several relatively independent dimensions that distinguish quite separate problem areas -- technical knowledge, level of training, frequency of use, and job definition, for example -- that may require different approaches to effective interface design. Useful discussions of these distinctions and their implications may be found in Cuff (1980) and Ramsey and Atwood (1979). In an ARI project currently in progress, Synectics (1980) has set forth criteria for selecting dialogue techniques based on task and user characteristics.

Adapting some observations of Cuff (1980), we can summarize the characteristics of the "casual user" as including:

- Poor retention of detail. Since they use the system only occasionally, casual users tend to forget both major and minor details of it. Knowledge of the system is not subject to the constant reinforcement of practice.
- Propensity for error. Sometimes users are aware that they have forgotten (or never knew) some operational detail, but they try a reasonable guess. On other occasions, they simply misremember, and confidently enter the wrong input. The possibility for error is particularly high when the user has to remember and select from a set of alternative actions without help from the system.
- Need for a safety net. Inexperienced casual users, especially, expect to make errors. They expect to find their way eventually, but they also expect the system to catch them when they stumble. They need to feel that they will never be left in limbo, not knowing what to do in order to get back to some recognizable place.

- Limited typing ability. Most casual users have only rudimentary typing skills, and they are unlikely to want to undergo training to improve their familiarity with the usual keyboard.

- Limited initial training. This characteristic is quite variable, of course, but as a general rule, any system interface procedures that require a lengthy training period to learn will be difficult to remember adequately without constant reinforcement.

- Reluctance to use documentation. Casual users are intolerant of systems that require them to search through a manual to determine an appropriate input or to interpret an output.

- Intolerance of structural formality. Casual users, by and large, are not acquainted with the precepts of formal logic or data base organization, and are frustrated with interfaces that are legalistic or that require them to be formally precise. They expect the system to be able to interpret what they mean, using at least a minimal level of common sense.

A continuing challenge to system developers is that of designing systems that meet the needs of both novice and expert users. System characteristics that may be appealing and useful to a novice may be the cause of frustration and inefficiencies for the expert. This fact has several implications. First, it means that system evaluations based on use either exclusively by novices or exclusively by experts can lead to conclusions that will not generalize beyond the type of user with whom they were obtained. Second, it represents a challenge to system designers to design into the system the kinds of flexibility that will accommodate users of all levels of expertise. Third, it represents a special case of a more general issue in the design of interactive systems, namely the desirability of providing those systems with the capability of adapting not only to the skill level but to the idiosyncrasies of individual users.

3.3 User "Styles"

A notion that one encounters among computer users is that of user "styles." The idea is that different users have different preferred ways of interacting with a system. The idea is related to that of "cognitive style" that one encounters sometimes in the cognitive psychology literature. Research questions relating to this notion include the following: Are there in fact identifiable user styles? Do these styles relate in a straightforward way to different prototypical ways of approaching intellectually demanding problems? Can a style taxonomy be developed that would provide some useful guidance to designers of interactive systems?

3.4 Psychological Barriers to Computer Use

The claim is sometimes made that certain types of interfaces are more acceptable to people in certain positions than others. It is believed by some, for example, that people in authority often object to hands-on use of computers on principle. The notion, which is sometimes referred to as "pride of rank," is that one's "station" in life prescribes what is and is not viewed as appropriate work.

To the extent that this view is held in the military it could have some unfortunate consequences regarding the ability of people to interact effectively with computer systems. For some types of applications, "hands-on" usage is probably much more effective than usage through an intermediary.

Research directed at understanding better the "pride of rank" notion and its implications for the use of interactive systems could be useful both in discovering ways to break down such prejudices and/or to design system interfaces so as to be compatible with prejudices that are difficult or impossible to modify.

Chapanis (in press) touches on the fact that many people seem to have fairly strong attitudes toward computers. Some of these attitudes are positive and some are negative; what is common about them is the fact that they typically are strong. The topic of attitudes raises two questions that seem particularly germane to the purposes of this report:

- To what extent are some of the popular aspirations and fears regarding computers well founded?
- To the extent that strong positive or negative attitudes are <u>not</u> well founded, what role, if any, should human factors specialists play in attempting to dispel them?

3.5 User Acceptance of Innovation

User acceptance is always a problem that must be faced when introducing innovative and, in particular, computer-based systems or procedures into a situation in which there are well-defined ways of doing things. Studies that would predict the kinds of user resistance that would be likely to be encountered and how these problems could be defused would be useful.

The willingness of an individual to accept an innovation, such as an automated decision aid, may depend to some degree on whether he has played a role in designing it. Moreover, there is some evidence that, as a general rule, people may feel better about and perform better in situations over which they have some control (or at least believe themselves to have some control) than in those over which they do not. Monty, Perlmuter, and their colleagues, for example, have conducted a series of studies the results of which indicate that students perform better on tasks they <u>believe</u> they have played some role in choosing than on tasks that they perceive to have been imposed on them (e.g., Monty, Geller, Savage, & Perlmuter, 1979; Perlmuter, Scharff, Karsh, & Monty, 1980).

3.6 Determination of Job Requirements

As we have already noted, the introduction of computer-based tools into work situations can complicate the problem of job/task specification. The complication arises from the fact that one of the effects of the introduction of such tools is that of changing qualitatively the nature of the job the user performs. If one overlooks this fact when specifying jobs in anticipation of introducing new tools to facilitate the performance of those jobs, one may end up specifying jobs that will soon be obsolete.

3.7 Skill Maintenance

Significant problems relating to the operation of some computer-based systems are those of boredom and job dignity. Skill maintenance may be an especially difficult problem in jobs requiring only the monitoring of an automated process. The difficulty stems in part from the great disparity between what the person must do when the system is functioning properly, which hopefully is most of the time, and what he must do when it malfunctions. When the system is functioning properly, the person may feel not only bored but useless. When it malfunctions, it is essential that he act quickly and skillfully; but if the day-to-day operation permits him to lose interest in the job and presents no real challenge, how is he to maintain the level of skill that is necessary to perform effectively in those emergency situations that may arise? There is a need for research addressed to the problem of the maintenance of skills that are infrequently exercised.

3.8 Safeguards Against User Error

One of the frequently mentioned desirable properties of a "friendly" system is tolerance of user error. The notion is that the system should be given the ability to detect user errors and to correct them, or at least to flag them so the user can take corrective action. To program an error-detecting ability implies knowledge on the part of the system developer of the kinds of errors that people are inclined to make. One must assume that many errors are system-specific, and for such errors a taxonomy would have to be developed for the specific systems involved. A question for research is whether there are certain types of errors that tend to be made independently of the system or at least that are made across a variety of systems. To the extent that such common error types could be identified, a firm basis could be provided for an effort to develop procedures for identifying and/or correcting them automatically.

It is customary these days to build into computer-based systems certain safeguards against errors that would have especially undesirable consequences. For example, before executing a command to delete a file the system might ask the user for an explicit confirmation that the delete command was what he intended. In some cases, a second level of protection is provided such that even after the system accepts the confirmation of the delete command, it defers the actual execution of the delete function until the end of the work session.

While the general notion of building safeguards against user-initiated catastrophic errors seems a very reasonable one, the question does arise as to how safe is safe enough. Imagine a system that presents one with the following sequence of queries before carrying out an instruction that could possibly be in error: "Are you sure you want to do this?" "Yes." "Are you <u>really</u> sure?" "Yes." "Last chance to change your mind ...." At some point the user's response would be likely to become violent. We are not suggesting that real systems have provided this degree of protection of users from themselves; however, the issue is a real one, especially as it relates to the problem of designing systems that meet the needs of users of all levels of expertise. The degree of protectiveness that may be appropriate for a novice user may be perceived as patronizing or obstructive by an expert. Again, the point arises as to the desirability of designing systems that can be tailored to the users' idiosyncratic needs. The following rules of thumb are sometimes followed in building safeguards into systems:

- Insofar as possible, no user should be given the ability to do something that is catastrophic for another user.
- Safeguards designed to protect users from their own actions should be designed with novice users in mind.
- It should be possible for experienced users to inactivate certain specific safeguard features if they choose to do so.

Input reliability is a well recognized problem in military information systems. A question for research is how to increase that reliability. There are certain obvious things that can be done. For example, the user should always be given feedback regarding what the system has interpreted his input to be. For inputs the accuracy of which is particularly critical, it may improve matters to have the user enter them twice. In some cases, it may be possible to check an input for consistency with other information already in the data base.

There undoubtedly are some procedures and policies that are used in other contexts to maximize the reliability of data input that would be applicable in the tactical situation as well. One suspects, however, that there is a need for the development for new ways of dealing with this problem, and that is a task for research.

3.9 Organizational Impact of Computer-based Systems

The tools that people use to get work done sometimes have organizational implications. Some tools will not fit readily in some organizations. In such cases the attempt to introduce such tools may force organizational changes. If the necessary organizational changes are resisted, the tools may not work effectively.

The introduction of increasingly sophisticated aspects of information technology in military contexts undoubtedly will have some organizational impacts. What these impacts will be may not be easy to anticipate. However, the effort to anticipate them through appropriate research should decrease the probability of undesirable effects of changes that are forced and unplanned.

4. INTERFACE ISSUES

The interface between the user of a system and the system itself has typically been the place where human factors work on user-computer systems has focused. This is as it should be, because the interface is, by definition, where the coupling between the user and the computer occurs and the place at which impedance mismatches will have their most obvious effects. A limitation of the work that has been done on interface design, however, is the lack of attention that has been given to the more cognitive aspects of the problem. One of the implications of the increasing pervasiveness of computer-based tools is that the cognitive aspects of the problem of matching systems to their users will become increasingly important. Indeed, the term "cognitive interface" is probably a useful one to characterize the types of issues with which system builders will, more and more, have to deal.

In the following pages, we discuss various aspects of the user-computer interface, with particular emphasis on cognitive issues. To structure the discussion, we adopt a viewpoint somewhat analogous to that of communications technology, a discipline that also faces the problem of information exchange between unlike systems.

The communications engineering community has developed a layered model to deal with the several facets of intersystem communication in orderly fashion.** The model's layers ascend from the physical media that support the transport of information through the protocols that govern the use of those media to the applications served by the communication process. The full model contains multiple layers, but for our purposes it will suffice to think of the three layers suggested above.

In user-computer communication, the physical layer can be thought of as coupling the input/output subsystems of a human (the auditory, visual, tactile, vocal, and motor systems) with those of a computer (display and printing equipment, motion or position sensing devices, and acoustic input/output equipment). A critical characteristic of any such coupling is its bandwidth -- a measure of the speed at which information can be transferred across it. Bandwidth, in turn, depends on the physical capabilities of the respective input/output systems, and the degree to which they are matched to each other.

We believe that within the limits established by the current state of technology, much remains to be done to increase

---

**The model, known as the Open Systems Interconnection Model is described fully in ISO, 1978 and in more tutorial fashion in ISO/TC97/SC16 N117 November 1978 Open Systems Interconnection.

bandwidth and to improve the communication across the user-computer interface. In pursuing that goal, what we are doing, in effect, is tuning the computer's input-output equipment so as to minimize any mismatch between machine and user.

Computer display technology flourished in the early to mid 1960s, but then lay dormant through the late 1960s and the first half of the 1970s, limited in applicability by its high cost. Alphanumeric display terminals were widely employed, but systems with a full graphic capability were limited to research laboratories and the relatively few applications that could pay the price. The rapidly decreasing cost of computer hardware has by now brought down the price of computer graphics to the point of economic feasibility for many applications, and that trend is expected to continue for a number of years to come.

The enormous potential of graphic display technology for effective user-computer communication has yet to be more than partially realized. Whereas in the 1960s the high cost of graphic systems limited the exploration of their human interface potential, in the late 1970s costs came down so fast as to outstrip the ability of human factors researchers to keep up with hardware developments. Thus, we now have something of a research vacuum in terms of the human interface applications for computer graphics. Section 4.1 addresses what we consider to be some of the principal research directions for harnessing that potential.

Speech processing technology presents a similar picture: there is a substantial but largely untapped potential for high bandwidth communication between user and machine; advancing technology and descending hardware costs make the exploitation of this technology an increasingly attractive proposition. Speech input and "understanding" by computer systems and speech generation are both discussed in Section 4.2.

Graphics and speech technology represent promising, but not the only, input/output topics for human factors research. In Section 4.3, we discuss other approaches toward widening the user-computer bandwidth. These include multi-media communication techniques that may involve speech, graphics and other media, eye fixation tracking as an input method, and other types of high bandwidth input.

In the user-computer context, the word "interaction" implies a running exchange of information or dialogue between user and computer. The need for such a dialogue was recognized from the beginning of interactive computing (Licklider, 1960) and virtually every interactive system has supported some form of user-computer dialogue. In communication engineering terminology the user-computer dialogue can be thought of as a protocol that defines the format, syntax, and time sequence of information exchange between two systems. Protocols assume the existence of an appropriate physical medium for information transfer; they are an absolute necessity if meaningful information exchange is to take place through that medium. In Section 4.4, we discuss the design of user-computer dialogue. Where possible, the discussion avoids hardware-specific issues, but where some aspect of a dialogue form depends on particular hardware, that dependence is made explicit.

The final layer in the communication model represents the application served by the process of information exchange. At this level, one can conceptualize the interface between a computer system and its user as an interface between two information structures. One information structure resides in the computer, and the other in the user's head. Presumably, as a result of the interaction, one or both of these structures will be modified.

At this level, our primary concern is with the semantic content of these structures and the operations that affect the structures. In some sense, we would like to establish a good "understanding" between user and machine. As is the case with input and output, we have no control over the design of the user's mental abilities and only moderate control over the way those abilities are employed. Consequently, we must again focus on the machine side of the equation. At this level, however, fewer constraints are imposed by fundamental limits in the technology. We may need sizable amounts of storage and processing capacity, but these are becoming increasingly affordable with the passage of time. Instead, the main concern is with the design of software so as to create computer information structures compatible with those of the user. Various approaches toward this goal are considered in Section 4.5.

In this section, we touch on the matter of user models of a system from the viewpoint of the software designer. The structures for storing information and the operations made available for locating the information stored in these structures form an important part of any user model. This is a particularly important issue given the fact that many military systems have as their primary role the storage and retrieval of information.

Although a well-designed system will present an intrinsically understandable model to its user, the process of understanding can be aided by effective documentation and by user aids built into the system. These topics are also discussed in Section 4.5.

The foregoing sections tacitly assume that although a computer system may be made understandable and "friendly" to its user, the information structures of the two will remain fundamentally different. For two decades, artificial intelligence researchers have pursued the goal of creating computer information structures equivalent to those of the human mind. Until recently, the results of this research were limited to laboratory demonstrations. As in other areas of the technology, descending costs are now making it possible to apply these techniques in practical systems. In Section 4.5, we also discuss some research questions opened up by these knowledge-based tools.

4.1 Displays

Because they have no moving parts, and can couple closely to computer memory, electronic displays represent one of the highest-speed computer output devices. Because they couple directly to the human visual system, such displays offer the potential of a very high bandwidth channel for information transfer from computer to user. The principal human factors task with regard to displays is to realize that potential to the maximum extent possible. If coupled to appropriate sensor devices, displays can also become part of a two-way channel between user and computer. The use of such channels to support user-computer dialogue forms is discussed later. Here we confine the discussion to the use of displays as a computer output medium.

In the pages that follow, we touch on three principal issues that appear to offer the greatest human factors challenge for the years ahead:

- o the issue of graphic encoding forms for displayed information,
- o the use of displays to access very large quantities of information, and
- o the use of displays to present dynamically changing information.

These research questions focus on the manipulation and presentation of information through appropriate software techniques. However, it is necessary to consider briefly some hardware issues before proceeding.

4.1.1 A Note on Display Hardware

At the present time, we have available two principal types of computer display, random-scan (vector) displays and raster-scan displays, both of which depend on cathode ray tube technology. The differences between these displays have implications for the information presentation techniques that can be supported, and so these differences must be considered by the human factors specialist.

Vector displays trace out an image, calligraphic fashion, by moving the display's electron beam in some pattern of straight or curved strokes. The pattern for a complete image is represented by a list of display commands. The display system contains special-purpose hardware, often quite elaborate, that converts the display list to strokes on the screen. Typically, the display list will represent an image in compact form, with each command specifying a straight or curved line, character, or geometric shape. Groups of commands are employed subroutine fashion to represent repeated picture elements. The supporting computer creates or changes displayed images through appropriate manipulation of the display list.

Raster displays sweep the electron beam in a constant fixed pattern similar to that of a television receiver. "Bit-map" raster displays make use of a dedicated memory such that one memory cell controls the beam intensity (and/or color) at each point on the screen. To create or change an image, the supporting computer places appropriate values in the desired memory cells. The memory serves, in effect, as a point-by-point map of the image.

Vector and raster displays have fundamentally different properties in terms of the images they can support. State-of-the-art raster displays can display a million points with as many as 2^{24} color/intensity values per point. They also can present both line and tone graphics, including realistic, shaded half-tone renderings of solid objects.

Vector displays have a much lower presentation capacity. (If the display list grows too long, flicker will occur.) Vector displays are limited to line presentations; they cannot, in general, present areas of tone value. On the other hand, a vector image can be constructed and changed rapidly -- by establishing and editing the relatively compact display list. In contrast, a bit-map image takes much longer to construct, since the value of each point must be computed; and, except for minor changes, it takes longer to alter a bit-map image. Thus, in general, vector displays excel at presenting line images, including dynamically changing images, whereas bit-map displays excel at presenting rich, tonal images, but ones which exhibit only limited degrees of rapid change. There is much work in progress on new display technologies, including flat-panel plasma or liquid crystal displays and large-screen projection displays. However, most of these displays exhibit properties similar to those of either the random-scan or raster-scan displays mentioned above. For further discussion of current display technologies, see Nickerson et al (1980).

4.1.2 Coding Parameters for Dynamic Displays

The problem of how to code and represent information in displays has received a great deal of attention from human factors specialists. Computer-driven displays, with their immense versatility, add new dimensions to the problem. They provide the possibility of coding information and highlighting various parts of a display in a great variety of ways, e.g., differential intensification, blinking, color changes, use of moving cursors, spotlights, windowing, zooming in, zooming out, selected display of Boolean combinations of data base components, exploding information on demand, and so on. Effective exploitation of the full versatility that computer-driven displays make possible is a challenge to graphics programmers and human factors specialists alike.

One issue in coding is the choice of symbols. With computer generated displays, the types of symbols that one can use are, for practical purposes, unlimited. There is a need for research, however, to identify symbols that are optimal in specific contexts.

The problem of developing guidelines for display formatting is complicated by the fact that different users of a computer network may have terminals with different capabilities and characteristics. The sender and receiver of a message through a computer-based message system, for example, may have different types of terminals. The difference may be as great as one person having a paper-output device and the other a graphics display. A format that would be suitable for one of these terminals may not be suitable for the other.

Sometimes familiar problems emerge in a new context and the context may have implications for the way the problem is approached. Consider, for example, the following problem that is conceptually straightforward and a relatively common one to human factors specialists, namely that of selecting a set of colors that are to be used in a multi-color display. The problem is as follows: A specific graphics terminal displays color by combining three primaries at each pixel. The color of the pixel is determined by the relative intensities of each of the primaries. The display hardware permits 2^8 or 512 intensities on each primary, from which it follows that each pixel is capable of displaying 2^{24} or roughly 17 million different colors. Only eight of these colors, however, can be used by any given program. The display software specifies for each pixel one of the eight preselected colors with a three-bit code. This three-bit code is replaced with one of eight 24-bit codes to activate the display. The human factors problem in this case is that of selecting eight colors from the 17 million possibilities.

To put the problem in a specific context, consider the following application: A graphics program has been written to display, singly or in combination, various layers of material in an integrated circuit. Conventionally, different circuit materials have been color-coded according to the following scheme: green to represent diffusion; yellow, ion implantation;

red, polysilicon; blue, metal; and black, contact cuts. The problem is that one wants to choose colors that will be recognized by their names and also will be maximally discriminable one from the other. In addition, one wants other colors that will be perceived as combinations of these basic colors and still maximally discriminable from each other. This specific application of color graphics is not one that is likely to be used by a tactical computer system, but the coding problem it represents has analogs in many uses of multicolored displays.

4.1.3 Information Access

A typical alphanumeric computer terminal can display about 24 lines of text, each up to 80 characters long -- somewhat more than one-third of a single-spaced typewritten page. A state-of-the-art high resolution display can present the equivalent of a full single-spaced, typewritten page. A high resolution, bit-map display can present about one million points, approximately the same information as a relatively crude 10-by-10 inch drawing, map, or photograph (with 100 lines per inch resolution) or a more refined 5-by-5 inch example (with 200 lines per inch resolution). A 30-inch by 60-inch desk can easily display to its user a dozen or more typewritten pages or 10-by-10 inch drawings -- the equivalent of eighteen state-of-the-art high resolution displays. Moreover, the desk's user can rapidly scan and easily manipulate the large quantity of information presented by his desk, employing techniques so natural to him as to require little, if any, conscious effort.

In contrast, the user of a display suffers from a serious handicap. It is no trouble at all to put a dozen, or several thousand, typewritten pages (or drawings) behind the display -- that is, in the computer. However, the display provides a very narrow view indeed of what is behind it. Its user is in a position not unlike that of a submarine commander whose view of the world is limited by the optics of his periscope.

There are two fundamental approaches toward overcoming this limit of display technology. One is a "hard" approach: increase the presentation power of displays by increasing their size and resolution. Such fundamental advances in display technology are the object of active research. Plasma displays, solid state flat panel displays, and CRT displays are all under active and continuing development.

The "soft" approach toward improving display usefulness focuses on the fundamental advantage that a computer-driven display enjoys over a desk or other conventional presentation device. That advantage lies in the fact that what is shown by a display can be changed at electronic speed by the computer that drives it. To make today's display technology more useful, we must find ways to take advantage of this capacity for rapid change.

Heretofore, processing and storage capacity have been the limiting factors in exploiting the dynamic properties of displays. Limited forms of change can be accomplished easily, such as the ability to "scroll" a text file past the "window" of an alphanumeric display. However, to provide the researcher or system designer with real freedom requires significant dedicated processing power and high capacity frame buffers, resources that until recently have been prohibitively expensive for many uses.

At the present time, however, the costs for computing and storage hardware are descending rapidly (25 to 40 percent per year) and this is expected to continue for the foreseeable future. This trend will very soon make it economically feasible to dedicate very substantial hardware resources (the equivalent of a present day mainframe computer) to the single user, which is what is needed to exploit the dynamic potential of computer driven displays.

Relatively little basic research has been done on how best to make displays useful as data presentation devices, but displays will become increasingly important as interfaces in user-computer systems, so this is a subject that is ripe for research. The basic question is, given a substantial amount of data to be viewed and manipulated, how best one can overcome the fundamental limit of display systems, namely, the fact that they can present but a small fraction of the data at one time.

What makes the question interesting, and difficult, is that the conventional storage and presentation devices present such formidable competition to the designer of an equivalent computer information system. It is difficult to imagine interacting with a computer in the near future with anything like the ease and fluidity with which one can scan the surface of a desk or manipulate the papers that rest on it. On the other hand, the designer of an information system has available a range of possible techniques whose creative potential has scarcely been touched by all the research on this problem to date.

An examination of existing work suggests three issues to consider: what images to show the user, how to change those images so as to present over time all that the user wishes to see, and what set of manipulations to provide the user for "driving" the changing display.

All three issues present challenges, as is apparent from even a cursory review of existing work. With the expectation that in the foreseeable future computer systems will have effectively unlimited storage capacity, one approach is to think of a CRT as a window on a "virtual" display of immense proportions and minute detail. Consider the following fantasy: Suppose it were possible to store within the computer a model of the globe with detail at a level that would permit the inspection of a map showing the locations of individual houses on a street. Now suppose one had the means to inspect any part of that representation at any desired degree of resolution. One can imagine viewing the entire globe (or at least a two-dimensional view of one hemisphere) and then being able to zoom in on a continent, a country, a region, a city, a city block, and then to zoom back out, rotate the image, and zoom back in on another area at whatever level of specificity is desired.

Realization of this fantasy is far beyond the current state of the art, but some approximation to it is certainly possible. Consideration of such a capability raises a number of questions about how best to match display technology to human capabilities and limitations. For example, when one must examine a complex pattern by looking at small parts of it through a "window," does one get a better representation of the total pattern if one moves the window over the fixed pattern or if one moves the pattern under a fixed window? Note that the use of a CRT as a window is more analogous to the latter case; i.e., in panning with a display it is as though the pattern is made to move under the display surface.

When viewing a segment of a large display, how important is it for the viewer to know the location of what he is looking at with respect to the larger context? One possible way to present this information would be with an inset, say a small rectangle, representing the virtual display with a much smaller rectangular cursor on it representing the location of the CRT window with respect to the boundaries of the virtual display area. When the window is slewed the viewer would see the virtual display moving beneath the window and at the same time he would see in the inset the cursor moving on the vi:tual display surface. Note that if one wanted to move the window to the east (right) one would see the virtual display moving beneath the window from east to west (to the left) while the cursor on the inset moved east.

In building up a composite mental image of a virtual display, how important is it to maintain continuity in viewing that display through a window. Consider the following specific question: Suppose one is going to try to build up this image by viewing successive segments of the virtual display. What are the advantages or disadvantages of viewing those segments discretely as opposed to showing the observer not only the segments but the slewing from one to the next? How advantageous (or disadvantageous) would it be to add to the zooming capability so that one could look at one segment, then zoom back and see the entire virtual display at a lower level of resolution, then zoom in on the next segment?

Although much can be done to overcome the physical boundaries of a CRT screen, a return to the contrast between conventional and electronic information media suggests just how much remains to be done. With a desk, it is natural and easy to bring new parts of the total information into focus by picking them up and moving them, or by moving oneself -- actions so natural that they in no way impede the flow of work. The gestures required to pan or slew a displayed image require considerable practice, though in time they can become quite natural. However, when combined with the loss of peripheral vision, it seems doubtful that this way of manipulating a display system can compete with the manipulation of material on a conventional desk.

A second problem arises from the "flat" arrangement of information as conventionally presented by a computer display. A desk surface can hold and present far more than the number of pages mentioned earlier, simply because sheets of paper can be piled up in ways that leave them recognizable and easily retrievable by their users. Obviously, there are many questions here for research.

4.1.4 Dynamic Displays

An important aspect of military situations is the way in which they change in time. Moreover, the prediction of future changes can sometimes be made on the basis of an analysis of past trends. The use of computer-driven displays makes possible the highlighting of trends through the selective displaying of specified subsets of a data base and time-compression (showing successive time slices of a representation in faster than real time).

The effective use of computer-driven displays to represent changing situations requires a suitably designed language in terms of which the user can describe to the computer the displays he wishes to see and the various operations that can be performed on the data that are to be displayed, a topic considered in Section 4.4.

4.1.5 Content-format Issues Regarding Information I/O

The question of how inputs to a tactical information system should be structured or formatted is an important one. If they are to be highly structured, the computer can force the structure by prescribing the format explicitly to the user when he makes his input. It can, for example, present him with a pre-formed display, letting him fill in the blanks. This approach has the advantage that the information that gets into the data base is structured at the time of input and can therefore presumably be readily incorporated into the existing data base. It also has the advantage that the form that is presented to the user when he makes his input serves as a memory aid, reminding him explicitly of the types of information that are required and the forms in which they are needed. It also has the disadvantage, however, of constraining the user so that if the information he wishes to report does not readily fit a prescribed form, it may be difficult or impossible for him to report that information.

4.1.6 Computer-controlled Maps

A digital map on a computer-driven display can be used as an input or information gathering tool as well as for the purpose of providing information to the user. One can imagine, for example, a map displayed on a touch-sensitive panel so the user could input location information by pointing e.g., "I am located here" (pointing at a specific location on the map) or "The last sighting of unit X was here." To check whether it had obtained the information the user intended, the computer would display a blinking symbol at the appropriate spot.

In order to orient themselves on a map, many people find it convenient, if not necessary, to rotate the map so its orientation will be consistent with that of their body. The question of how important it is, in terms of error rate, to be able to rotate a map that is displayed on a graphics terminal, is one example of a human factors issue relating to the use of computer generated maps.

4.2 Speech

Speech as computer input is usually thought of in connection with natural language capability. In fact, it is a nearly orthogonal possibility. One could use speech as an input mode (e.g., using isolated words and a small vocabulary) without a natural language capability; and one can have a natural language capability (e.g., for unconstrained typed input) without speech.

Practical speech recognition systems for isolated words separated by pauses of about 250 msec have been developed and several groups are working on the problem of computer understanding of connected discourse. Because these developments are relatively new and the recognition of connected discourse has not yet reached the state of being operationally useful, little attention has been devoted to the human factors design of dialogues that exploit human speech. While it is acknowledged that speech can produce substantial improvements in speed of input, several questions must be answered before we know the circumstances under which such systems will be practical and before recommendations for interactive speech dialogue can be made. Can users easily adapt their speech to the requirement for artificial word segmentation? Can an individual sustain voice input in this mode for long periods of time? What are the most effective means of providing feedback concerning the correctness of encoding of the speech input? For what classes of data input is it most effective: numerical data? alphanumeric codes? text? If text is feasible, how does one integrate formatting and punctuation instructions with the text itself? What are the conditions under which voice is a useful mode to be preferred to CRT or hardcopy printout? These are uncharted areas in need of substantive research.

The technology exists to make feasible the development of computer-based message systems with the capability of handling voice messages, at least to a degree, although the costs would probably still be prohibitive. It would be technically feasible now to record and transmit voice messages while obtaining header information (addressee, etc.) with keyboard entry. It should be technically feasible soon to enter the header information by voice, although this is considerably more difficult than simply recording and transmitting the message, because it requires some speech recognition capability.

The assumption usually is made that voice input for a computer system is generally a desirable thing. The validity of this assumption is not beyond question, however, and whether, on balance, voice would be a preferred input mode for computer-based message systems is an open question. It seems very possible that voice will be a preferred mode for some applications and keyboard entry will be a preferred mode for others.

Obvious advantages of digitally-encoded voice systems include easy encryption and the naturalness of speech as means of human communication. Human factors problems associated with the use of such systems include intelligibility, quality, and speaker identification and authentication. All of these problems arise from the fact that bandwidth compression techniques destroy many of the normal auditory cues to speaker identity.

4.3 Other High-bandwidth Input Methods

As we have already noted, a general problem relating to user-computer interaction is that of how to increase the input bandwidth. A severe limitation of a keyboard input device is the bandwidth limitation it represents. As noted above, an obvious way of increasing the bandwidth that has received considerable attention from researchers for several years is to provide a speech recognition capability. There are other possible ways of increasing bandwidth, however, that have received much less attention and that deserve consideration. An important question for research is that of how to design interfaces that permit the user, even without using speech, to convey information to the computer at rates comparable to those at which he conveys information to other people.

4.4 User-Computer Dialogue

From a human factors point of view, the nature of the dialogue between user and computer is one of the most important aspects of a system's design. A well-designed dialogue form will promote efficient use of the system it supports and help clarify the user's understanding of that system. A poor design will hold back the user and confuse his understanding.

Among factors that an effective dialogue design must take into account are the communication medium, the available work station hardware, the system's functionality, the expected use pattern (routine, repetitive vs. varied, unpredictable), and the type of user (expert vs. novice, frequent vs. occasional).

Over the last two decades, interactive computing has been studied from a number of points of view. Nickerson (1976) cites several efforts to develop systems that permit dialogues that resemble conversations between humans at least in certain respects. However, in spite of the interest shown in this topic, there have been few attempts to study the subject in a manner sufficiently comprehensive to permit the development of definitive design principles. Given the growing importance of user-computer dialogue, this would seem to be a critically important research area.

The following paragraphs present a brief taxonomy of several basic dialogue forms that are available for the design of interactive systems. This material is taken from an earlier BBN report (Nickerson, Adams, Pew, Swets, Fidell, Feehrer, Yntema, & Green, 1977). In describing the alternative forms, we comment on their principal characteristics with regard to medium, hardware, system functionality, use pattern, and the types of users for which they are suited.***

4.4.1 Menus

With a menu structure, a set of alternatives is presented and the user is provided with a way to select one or more of the alternatives displayed. The response mode may involve moving a cursor, typing a symbol, or pointing with a light pen, stylus, or finger. Frequently, a hierarchical sequence of menus is provided to elaborate a tree structure of possible choices. A user who is naive with respect to computers is still likely to understand a menu procedure. The only requirement is that the terminology used to describe the choices be understandable. Whenever more

***The emphasis here is on dialogue form (e.g., menu selection vs. command techniques) as opposed to the supporting medium (e.g., spoken vs. typed entry of menu selections or commands). In other sections, we focus on the special characteristics of graphic and speech communication media.

than two or three choices are possible and the designer is not willing to assume that the user knows what the alternatives are, then a menu may be an appropriate structure to use.

4.4.2 Formatted Inputs

A form-filling frame is a computer translation of a printed form onto a CRT display. The extent and complexity of the form may vary from one designed to obtain a value for a single parameter or variable to one intended to support a complex data-entry operation. Generally, the fixed or background part of the form is a protected field, and is displayed at an intensity different from that of the data fields. The data fields may be of fixed or variable length and usually require a terminator. The data fields are editable, but the background fields are not. Only knowledge about the terminology in the background fields and understanding of rudimentary cursor control is required of the user to interact with such a form.

4.4.3 Question-and-Answer Inputs

Printing terminals are not suitable for form-filling frames, but an abbreviated question and answer format or prompting format can serve the same purpose. In these cases, a sequence of fixed or background fields is presented, one field at a time. The user fills in the appropriate response and is then presented with the next background field. If the background field is phrased as a question, it is called a question-and-answer frame. If it is more abbreviated, it might be called a prompted entry. In either case, the desired input may be a fixed-length code or a variable-length code, perhaps even an extended piece of free-form text that will be entered, but not interpreted.

In a typical message system designed for printing terminals, a prompted message composition sequence calls for entries in a series of header fields and then for the entry of the body of the message as a text field. In a CRT-based terminal, the same operations would be displayed as a form-filling frame.

Note that with any of the preceding methods, there is no ambiguity concerning what is required. Because the computer remains in control of the interaction, these techniques are particularly well adapted to inexperienced users.

4.4.4 Limited-Syntax Command Languages

In contrast to computer-controlled techniques are those in which control is transferred to the user at points where what is to be done next has intentionally been left incompletely specified. In this case, the user has the opportunity to initiate commands rather than simply to select a response from among those suggested. Illustrative of such techniques are limited-syntax command languages. Such languages are direct descendants of interactive computer programming languages in which commands may be either stored for future execution or executed immediately upon being specified. These languages require considerably more knowledge on the part of the user than does use of the dialogue types mentioned above. The acceptable commands at any point in a transaction must be recallable by the user, not just recognizable as in the previous cases.

By use of the term "limited syntax," we mean to suggest that the structures in which the commands must be formulated are predefined. Typically a command line begins with a verb such as "edit," "compute," "locate," etc. The verb is followed by one or more arguments that successively bracket the domain and specify the values of parameters to be addressed. When the command is parsed, the system expects to find one of a pre-defined set of possible arguments at each point in the sequence. Thus both the structure and vocabulary must be learned by the user. Advantages of limited-syntax structures are their conciseness and efficiency. If a particular activity is frequently repeated, it would be tedious and unacceptable to an experienced user to be forced through a series of menus or form-filling frames to define every desired activity.

In systems in which considerable training can be justified, command languages can be made still more concise by introducing abbreviations and brief codes for both command verbs and arguments, as has been done in commercial airlines reservation systems. This renders the transcript of a transaction relatively unintelligible to the uninitiated, but provides for substantially more rapid command entry.

Some command languages provide a middle ground between terse but demanding command languages and easy-to-use but verbose form-filling or prompted input techniques. At each point at which a new argument is to be introduced, the user has the option, in these systems, to depress a special key. Depression of the key causes the system to provide the user with a brief description of the class of argument that it is expecting at that point in command formulation. It should be noted that increasing the versatility of command structures generally requires much more extensive frontend software to make the dialogue flow smoothly.

4.4.5 Special-Purpose Function Keyboards

Another way to promote conciseness is to avoid the requirement for typing each command verb by using a labelled <u>special-purpose function keyboard</u> that identifies each command with a single key press. Such a keyboard has the advantage of providing a reminder regarding what commands are available at any point without the concomitant constraints of requiring the user to step through a lengthy menu or form-filling frame. Sometimes the special keys are integrated into a standard alphanumeric keyboard and sometimes they are designed as a separate block of keys for the designated purposes. Some special-purpose keyboards are provided with labelled overlays that permit the user to change the assignment of keys according to the requirements of different software packages, but to change assignments within a particular application can be confusing and can lead to high error rates. As a practical matter, overlays also have a tendency to get lost or misplaced, placing still other demands on the user's memory of key assignments.

An example of the effective application of a special function keyboard is provided in the simulation of a remotely-piloted vehicle (RPV) command and control system developed by the Air Force Human Engineering Division. This system employs IBM 2260 terminals dedicated to monitoring and controlling a flight of RPV's. Each of four CRT terminals displays flight path information and vehicle status information. It has a light pen, an alphanumeric keyboard and a separate function keyboard to activate frequently-used commands. These commands specify such operations as changing display scale factor, changing RPV speed or altitude, and modifying flight path. Many of the commands require an argument that is provided either through a light pen designation of the vehicle whose status is to be changed or a description of the revised flight path. The alphanumeric keyboard is used to enter new parameter values such as velocity or altitude. The special-purpose keyboard is particularly effective because users sometimes work under severe time constraints and typically need to issue well-defined and unchanging commands.

4.4.6 Natural Language

There has been considerable interest in developing the capability to implement user-computer dialogues in relatively free-form natural language. This capability has been pioneered in the development of computer-assisted instruction systems in which the users are dealing with subject matter that is rich in vocabulary and for which the unnatural constraint of menus or question-and-answer formats communicates the image of an instructor as a robot with a relatively limited ability to communicate. The problems involved in implementing natural language dialogues are many and well known. Such a capability requires a large stored vocabulary with an associative network of synonyms and related concepts, a parser of natural language grammar, a sentence-constructing algorithm for synthesizing answers to questions that arrive in unpredictable formats, a historical record of previous sentences, and their analysis in order to interpret the context sensitive features of human discourse, and so forth.

Sufficient progress has been made on these problems that it is now feasible to apply natural-language systems, in a limited way, in certain contexts where errors of understanding are unlikely to be disastrous; however, the technology is not sufficiently mature to be used very extensively in highly sensitive contexts such as tactical command and control. On the other hand, natural language interpreters will be important components of future developments in computer-based information and retrieval operations in which data storage and retrieval algorithms will be based on the semantic content of natural language messages rather than simple scans for key words or prestored phrases. Whether a fully natural language capability is a reasonable, or even desirable, goal for the foreseeable future is a debatable question. There can be no doubt, however, that much can be done to make user-computer dialogues more conversation-like than they currently are (Nickerson, 1976).

A researchable question of considerable importance for both dialogue development and information processing operations is whether individuals can learn to communicate using a constrained subset of natural language grammar and vocabulary that would reduce both the verbosity of natural language dialogue and the difficulty of computer interpretation of human input. Such a dialogue mode may be thought of as intermediate between limited-syntax command languages and natural language; it should provide further useful information for improved structuring of command languages for ease of human use.

Relatively little empirical work has been done on the effectiveness of such languages; however, one study of the effects of restrictions on vocabulary size was reported recently by Kelly and Chapanis (1977). Two people were placed in separate rooms and had to communicate by teletype. They worked on two tasks requiring interactive problem solving. The investigator compared performance with (1) an unlimited vocabulary, (2) a 500-word vocabulary consisting of 425 function words and 75 task-related words and (3) a 300-word vocabulary consisting of 225 function words and 75 task-related words. In each case, the subjects were given training with the admissible vocabulary. Time to solve the problem and total number of communication exchanges proved to be independent of vocabulary size in this study. Much remains to be done, however, to explore minimum vocabulary size, restrictions on grammatical constructions, and the generality of this finding to other classes of tasks.

A very specific research question that relates to use of computers in tactical situations is the following one: What would be a minimum-word vocabulary that would be adequate for describing tactical situations? This question probably does not have a single-number answer. It seems more likely to be the case that a vocabulary of a specific size would be adequate to describe a certain percentage of situations that could arise. If that is the case, what one would like to know is the shape of the curve that relates number of words to percentage of situations they can describe. Of course, this also has to depend on the specific words that are chosen, but therein lies a challenge for research.

4.4.7 Graphics Language Development

A question of some interest is whether one could develop a communication language that is suitable for use with a computer-based communication system with a graphics terminal that contained as language primitives words, icons (graphic symbols), and maps. A purpose for such a language would be to permit the user to describe tactical situations effectively. A research problem that follows from this question is the need to identify the capabilities that a graphics system (including its command language) must have in order to facilitate the representation of the dynamics of a changing situation.

4.4.8 Customizable Interfaces

The assumption is sometimes made that there is merit in designing an interface so as to provide the user the opportunity of customizing it to his own preferences or style. Providing such flexibility is not a trivial matter, however, and data apparently do not exist that would provide the basis for a judgment as to the real value of such flexibility from the user's point of view. There are some systems in existence that provide such flexibility. It might be useful to attempt to determine the extent to which users of such systems take advantage of this flexibility and, in fact, customize the systems to their own preferences. There is a chance that such a study would reveal that self-initiated changes are relatively rare, in which case one would have to question whether such flexibility is worth the cost of providing it.

4.5 Tools and Procedures

4.5.1 System Design and User Models

The user of an interactive system has in his mind a representation for that system. The representation begins to form on first exposure to the system; it grows and may change with subsequent use or instruction. In our experience, this representation, or "user model" as we shall call it, plays a pivotal role in the success or failure of a system from a human factors point of view.

To see the importance of user models, consider the difference between a conventional office and a computer system that automates some part of that office. The conventional office contains information, processing elements, and people. The information is represented in tangible, visible form on paper, which can be placed in folders, bound into reports, piled on desks, and stored in file cabinets. The totality of information storage in a conventional office may be confusing, but one would not normally characterize it as mysterious. The processing elements in a conventional office -- pencils, typewriters, adding machines, staplers -- are equally understandable to the human user. One may not be particularly adept at their use, but it is generally quite clear what they do and, in most cases, how they do it. This description stands in sharp contrast to an equivalent interactive computer-based office system. The latter may bring with it enormous advantages over the conventional office, but it can also bring great mystery. In the electronic office, processing and storage elements are reduced to microelectronic form, and this has the effect of making them intangible and invisible to the human user. The organization of a conventional office reveals itself through the physical realization of that office. The organization of an electronic office does not. In order to understand the electronic office, therefore, its user must compensate somehow for this lack of comprehensible physical reality. Most users appear to do this by constructing a mental model.

Once constructed, this model will guide the user in subsequent interaction with the system. Depending in large part on the success of the system's designer, the user will characterize his model as "clear," "easy to remember," "comfortable," or "confusing," "hard to keep in mind," "awkward." Individual parts of the total system -- its documentation, functions, objects, and command language -- will either clarify or obscure the user's model. The separate parts will reinforce each other, and therefore add clarity to the user model or oppose each other, and therefore add confusion. Taken as a whole, the user model will promote a friendly, willing user, or leave the user with a sense of distaste and a desire to avoid the system.

The foregoing comments have focused on the problem of automating an "office" rather than a military system designed for command and control, message processing, or some other military application. However, the contrast between the conventional and the electronic holds equally well in the case of military applications. In all cases, the problem faced is that automation makes the visible invisible and the tangible intangible. The problem is to compensate for the loss of physical objects, which, whatever their faults, have the useful properties of visibility and concreteness to the user.

In command and control situations, one must cope with the electronic replacement of maps, status boards, blackboards, and other easily understood and simply manipulated artifacts. In their ability to store, retrieve, display and compute data, the electronic replacements may have indisputable advantages over their conventional counterparts. However, unless this capacity can be made as easy to use, and to understand, as the older tools, most or all of the advantages will be lost.

At the present time, the shaping of user models can fairly be said to be an art rather than a science. It is guided by intuition, experience, trial and error. Our expectation is that it will remain an art for the foreseeable future. This arises from our perception of the "problem space" for user models as encompassing both the human mind and the totality of what can be done with computer systems.

The designer of an interactive system must know what he wants to put in the user's mind, what system characteristics will get it there, and how to build those characteristics into the system. To meet these needs requires an understanding of the psychology of human understanding, of available techniques for constructing the user-computer interface, and of the technology of interactive program design. In none of these areas do we have enough knowledge to provide the basis for exact, quantitative design disciplines.

However, this is not meant to discourage research on user models. Quite the opposite, our view is that this topic is an especially important one, and that it will become increasingly important as the systems that people use increase in complexity and potential mysteriousness. Moreover, the need to understand better how to give users appropriate and useful models of complex systems is doubly important, because when such models are not provided, the users are bound to develop their own models spontaneously, and the models they develop can often be not only inaccurate but also counterproductive if not dangerous.

With respect to the problem of providing suitable models of systems to beginning users, one may question the advisability of carrying over familiar concepts into the domain of computer-based systems. Consider, for example, the case of an electronic information storage and retrieval system. In the case of conventional data bases in which the storage medium is paper (e.g., books, journals, reports), one's memory representation of the information in one's own data base involves visual and spatial components. One may remember the color of the cover of a book or the design on the jacket, or one may remember the location of the pile in one's office in which a particular report may be found, or one may visualize a letter as being filed in a specific drawer of a specific filing cabinet. Such cues serve some useful functions. They help us locate material when we need it. They may also help us maintain a conceptual model of the way the information we have collected is organized.

In an electronic data base in which all the information is stored in computer files, such locational attributes of the information are lost. Should that loss be of any concern? Would one's use of a data base be facilitated as the result of the imposition on it of analogous secondary properties? For example, would it be helpful to have information dealing with different topics presented on different colored backgrounds on graphics terminals? Would it help to impose a spatial model on the information in the data base, thus creating the impression that different types of information are stored in different places? How important or useful is it for the user to have some model of the information that is in a storage and retrieval system and how that information is organized? To what extent would such a model facilitate both browsing and focused search?

As another example of how models that work with conventional information systems may be inappropriate in computerized systems, consider the case of data sharing. Computer science has developed information storage structures having properties that would be impractical or impossible to duplicate with conventional media. Such structures take advantage of the access speed and other properties of electronic storage media. They offer various benefits from a systems engineering point of view and also in terms of the functionality that can be achieved with an information system. From a human factors point of view, such structures present a challenge. If presented directly to the human user, they may create confusion. On the other hand, if exploited properly, they may help realize the potential advantages of an electronic information system as opposed to a conventional one.

Data sharing is a common technique in designing information systems. If a data element must be accessible from several locations in a total information structure, such access can be provided in either of two ways. One can place a copy of the data element at each location where access is desired, or one can provide just one master representation of the object, and then provide pointers, or references, from the access locations. The latter technique may prove to be preferable for any of several reasons. Two obvious ones are to save storage space (pointers may be much more compact than copies of an extensive data element) or to ensure consistency (with a changing data element, all references will automatically access the latest version).

Information structures built on the principal of copied data elements can be duplicated in conventional storage media. To make a certain memorandum accessible from several different files, a copy is usually placed in each file. When replicated in a computer system, such structures are therefore likely to be understandable to the human user. In contrast, it is not usually convenient to employ data sharing in paper-based information systems. Consequently, one would expect this technique to be less understandable to the inexperienced user of an electronic information system, and this has indeed proven to be the case. Confusion appears to creep in because the user believes (and is often led to believe) that he is accessing one of many copies rather than a single shared object.

4.5.2 Complexity

An interesting trade-off in designing interactive systems involves the need for complexity of a system on the one hand and a desire for simplicity and ease of use from the user's point of view on the other. This assertion leads immediately to a question: why complexity, and what is gained by it? What is there to trade-off? The answer lies in the desire for powerful systems, and, somewhat paradoxically, in the desire for conceptual simplicity and ease of use. The design of one computer-based message system, Hermes, yielded a number of experiences that may serve to illustrate this point (Myer, 1980). Hermes, like many other systems of comparable complexity, was not simply created once and for all, but rather it evolved over time. At the outset, the designers had little firm knowledge of what should go into a message system. An initial version was created with basic capabilities, and then interaction with the system's users guided its further development.

Very early, it became apparent that multiple options were wanted for the output of messages from the system. In some cases, one wanted a compact view of many messages. In other cases one wanted messages to be printed out, in their entirety, on the computer terminal. In still other cases, a listing was wanted of many messages on the centrally located high-speed printer of a computer center, and with a table of contents.

To provide these options, three separate output commands were included in the system's design. To provide further control over the output format for messages, a "Template" object was invented, and a specialized editor created, with its own set of commands to permit the user to create and modify templates. A new command was added when it became apparent that users wished to save partially completed draft messages from one session to the next. Another was devised to permit the saving of single message fields containing useful information, such as address lists.

Thus, as the system acquired power, it also acquired complexity. Unfortunately, interaction with the user community made it clear that complexity was not without its price. Some users elected to forego the power bought by Hermes' complex design in favor of simpler message systems offering fewer options.

In many situations, it may not be possible to switch to a simpler system. A wide variety of options may be necessary to get the job done, or there may be no simpler system to which to switch. Hence, we conclude that the question of complexity is a serious one in the design of interactive systems. The basic problem is how to provide the rich variety of options that may be needed or desired in an interactive system -- especially by experienced users -- without burdening the system with such complexity that it will be unattractive or even unusable to users with limited experience. We believe that this is a serious research question.

One line of attack is suggested by the observation that complexity appears to be a memory problem. Users seem to have trouble with complex systems in part because it is hard to remember how such systems work. This suggests a "get more from less" approach toward system design.

One way to do this is to provide only the most basic objects and most primitive operations, and then allow these to be concatenated in various ways to accomplish the end result desired by the user. The problem with this approach is that it may require unduly complex command sequences to specify even the most frequently used operations. All possible operations can be accomplished with such a system, but perhaps at the cost of an unacceptable level of difficulty of operation. The risk is that this approach may lead to a system which is easy to understand and remember, but hard to use, and frustratingly verbose for the expert user.

On the other hand, this approach has the power that many operations can be accomplished that were unanticipated by a system's designers, and without the need to add any new commands or objects to the system. Furthermore, frequent users of such a system often do develop a fluency in its command language that overcomes the more cumbersome constructs required. So far as we know, relatively few interactive systems have been built on this model. Hence, this may be a useful research thread to pursue.

Experience in the Hermes project suggests another approach that may be more practical, which is based on two policies: (1) Arrange the objects in the system into a small number of broad classes, such that all objects within a class are treated identically with respect to the operations that can be carried out on them. (2) Strive to create operations that apply to several, or all classes of objects, such that the result of applying an operation varies appropriately with the object class to which it is applied.

This approach is based in part on the characteristics of natural language and in part on the organization of physical objects that store information in conventional ways. In natural language, the verb "open" may have different meanings depending on the object to which it is applied. To open a file drawer is to unlatch and pull it out. To open a file folder is to unfold it. In a conventional office, a memorandum, a business letter, and a purchase requisition (whether filled out or not) are all objects of the class "document." The same operations -- writing, editing, filing, sending -- apply uniformly to all documents.

In a more recent message system design, this approach was applied in an effort to eliminate some of the apparent complexity of Hermes while retaining its power. In this system, templates, messages, draft messages, and fragments of messages are all instances of the object class "document," and are all subject to the same creation, storage, editing, and other operations. Also, in this design, single commands were made to apply where possible to multiple classes of objects.

The results of this effort were successful up to a point, but in the end it remained clear that command language design is an art, in which there is still much room for improvement.

4.5.3 Documentation and On-line User Aids

One of the advantages of computer-based interactive systems is the possibility of building into them the capability of providing their users with tutorial information about their use. Ideally, one would like such systems to have the capability of providing the kind of help to a user that is appropriate to his level of expertise. While many systems do provide on-line aids of various sorts, designing truly helpful helps has proven to be more difficult than it first appeared. (Nothing is more frustrating to a user than to type the "help" command on a system and then to discover that he needs help to decipher the response.) The design of these capabilities also remains very much an art form. Few empirically validated guidelines exist for developers of systems who wish to incorporate such capabilities.

4.5.4 Information Finding Techniques

Among the many purposes for which computer systems will be used by the military in the future is the storage and retrieval of large amounts of information. A key to the effective use of such information systems is the availability of tools that provide users with easy access to the specific information they need.

Information finding is, of course, a very general problem and is not unique to the military. Moreover, general purpose information systems are beginning to be developed for use by the general public. A study of these systems and, in particular, of the techniques they employ to provide user access to the data base, could be of considerable interest.

Several consumer-oriented information systems already exist, or are being developed. An example of such a system is computer-based information service provided to the general public by the British post office under the trade name Prestel. Prestel uses a specially adapted TV and a handheld numeric keypad. (The input is somewhat limited by the number of keys on the pad.) The communication link between the user's terminal and the computer in which the information is stored is a telephone line. Prestel's data base contains about 200,000 "pages" of information. A page is what is displayed on the TV screen at one time. The data base contains information on a wide assortment of subjects.

The user accesses the data base via a menu-structured query procedure. The computer presents a menu of up to 10 options from which the user selects one by entering its identification number with his keypad. The typical effect of the selection of an option is to bring up another menu on the video display from which another option is to be selected, and so on. The data base has a tree structure, so what the user is doing in selecting options from a sequence of menus is working his way down the tree so as to arrive at a terminal point of interest. Most of the information in the Prestel data base is entered and its storage is paid for by "information providers." The menu pages themselves are prepared by the British post office.

The kinds of human factors issues that arise with respect to the design and use of a system such as Prestel include: (1) the organization of information on the display, (2) the coding schemes used for keying queries or responses into the system, (3) the organization of the data base, and (4) the user's model of that organization.

Systems such as Prestel are intended to be used by anyone who has an interest in accessing the information they contain. They are to be distinguished from very special purpose systems designed for a highly restricted class of users, such as airline reservation systems, point-of-sale data entry systems, and inventory control systems. Inasmuch as the purpose of these systems is to provide a user with information that he wants, with a minimum of hassle, their effectiveness should be judged largely in terms of how difficult it is for the user to access that information. Short of being able to ask a question of the system in natural language, one wants to give the user the means of interacting with the data base in as natural and effective a way as possible. One wants to minimize the overhead involved (and the detours taken) in getting to the information that one really wants to have.

A common method of giving the user access to the data base is the one used by Prestel, namely that of providing the user with a sequence of "menus" by which he steps his way down selected branches of a "tree." One may question the efficiency of this approach from the user's point of view. If the user knows where he wants to end up, it seems unfortunate that he must be forced to traverse several branches of a tree in order to get there.

One of the problems that the designers of a data base have is that of matching the structure of that data base with the structure of information in the user's head. Assuming for the moment that people do think in terms of tree structures, the problem becomes that of organizing the data base tree so as to be similar to the user's tree. Evidence that this is not a simple problem comes from the fact that users often make mistakes in selecting from tree-searching menus (Frankhuizen and Vrins, 1980). They may know where they want to go, but not know how to get there in terms of the way the data base is organized.

The need for better techniques for interacting with (interrogating, browsing through) very large data bases is illustrated by some of the data bases that currently exist on the WWMCCS Information System. The Airfield Facilities File and the Units Status Reporting System, two applications software packages on WWMCCS, contain approximately 60 million and 75 million alphanumeric characters, respectively. The first of these data bases contains information regarding the physical properties and available facilities of approximately 47,000 air fields, and the latter contains information on the status, location, and capability of U.S. military resources throughout the world. These are not enormous data bases (a 400 or 500 page book probably contains on the order of one million characters, and so the Units Status Reporting System contains the equivalent of, say, 75 moderate-sized books), but they are large enough to illustrate the need for better data base search techniques. Moreover, the sizes of data bases will certainly continue to grow.

Imagine having electronic access to the Library of Congress, or even to the New York Times data base, which, as of 1975, contained nearly one billion abstracts of articles of its own and of other newspapers dating back to 1969. What might one do with it? Clearly not much, unless one also had available some useful interrogation procedures. But what constitutes a useful interrogation procedure? This is a disarmingly simple question, but one that does not have a simple answer. Some people anticipate the day when far-ranging questions can be asked (of the computer) and answered (by the computer) in free, unconstrained English (or other "natural" language). That day may come, but it is not here yet, and it is unlikely to come in the next few years. What, in the interim, is needed to make the idea of an <u>accessible</u> computerized general information store a practical reality?

4.5.5 Knowledge-based Dialogue Tools

As the capability-to-cost ratio of computer systems continues to increase, it becomes ever more practical to think of complex and sophisticated software in support of individual computer users. However, when one compares simple user-computer dialogues with human dialogues, a major difference is in the completeness and precision with which instructions for a computer must be specified. This completeness and precision requirement is one of the major hurdles for the unsophisticated user of computers. The human dialogue does not require such precision and completeness because the listener brings to the conversation a large amount of knowledge upon which he can draw to fill in the gaps and infer missing elements. The speaker tacitly assumes that this capacity is available and simplifies the message accordingly.

When one asks for the names of all the people who work for IBM and have PhDs, one does not mean the union of all IBMers and all PhDs. When one asks a secretary to type a dictated letter, one does not need to specify that the salutation goes on a separate line and the complimentary close goes on a separate line indented 15 spaces. When one asks, "What is the value of voltage at P-1?" and, after getting the answer from a technician, one then asks, "How about P-3?" or "How about the current?" one can expect that the context will allow the technician to give the voltage at P-3 or the current at P-1.

For several years the artificial intelligence community has been working toward computer-based representations of human cognitive processes, particularly as they are reflected in natural language. This work has the potential for automatically disambiguating incomplete or imprecise assertions of computer users.

Developing knowledge-based systems requires the close collaboration between the computer-scientist and the human factors specialist with knowledge of cognitive psychology and psycholinguistics. Such system capabilities can contribute major breakthroughs in the "friendliness" of interaction.

In addition to the problem of developing these capabilities, there are also researchable issues having to do with how to introduce such knowledge bases to the user. In the examples mentioned above, it is a simple matter to have the computer format the letter or have it report the intersection of IBMers and PhDs, although this always entails the possibility that the inference of what was intended was incorrect. In other cases, the issues are more complex. Imagine that the rules of entitlement to Social Security benefits or assignment of personnel to military specialties had been built into a knowledge base. One can then imagine computer support of the Claims Representative or personnel officer taking one of two extreme forms or some point on the continuum between them. In the case of the Social Security system, for example, the system could automatically and dynamically decide the most efficient question to ask at each point during an interview and arrive at the entitlement decision on the basis of the data entered in answer to these questions. Alternatively, the agent could conduct the interview according to individual style and the responsiveness of the interviewee so as to cause the interview to flow most naturally. Then, when all the data were entered, the computer would advise the agent concerning incomplete entries, missing data, tasks remaining, or, in the event there were no gaps, the decision. Probably neither of these extremes is best, but among the multitude of choices in between lie questions of display design, choice of interactive methods for prompting the agent, the nature and specificity of the prompting, and a host of issues that may be answered differently when such knowledge bases are available than when one is designing a simple frame-oriented branching dialogue structure.

5. SYSTEM ISSUES

5.1 System Architecture

Throughout the 1960s and well into the 1970s, interactive computing was dominated by a single architectural model: a central computer shared by multiple users, each supported by a simple terminal (one with no more than the most rudimentary processing or storage capabilities). This configuration was made necessary primarily by cost factors: computers were too expensive to dedicate to a single interactive user; resource sharing was necessary to reduce the cost of computing to an affordable level.

Throughout this same time period, however, two trends have been established that promise substantial architectural changes in the years ahead. The first of these trends relates to the costs of digital processing and storage technology. These costs have been dropping continually and dramatically, and at least two orders of magnitude of further decrease are foreseeable in the years ahead. At a low enough cost, it becomes possible to dedicate substantial processing and storage resources to the individual interactive user or to other functions that use those resources at less than peak efficiency.

A trend toward providing more computing power to the individual user shows up in the appearance of sophisticated terminals with built-in microcomputers and substantial memory, and in the incorporation of microprocessors into "smart" instruments and other devices not heretofore associated with computing. There now exist experimental systems in which as much computing power is dedicated to the single user as was heretofore incorporated into a large mainframe machine. Although the possibility of locating much of the computing power that a user needs close to, or in, his terminal is becoming economically feasible only now, the desirability of doing so was noted by Licklider as early as 1968:

> When connected to a large computer, a small one makes an excellent "intelligent terminal" or an excellent satellite. When so connected, the small computer can take care of the frequent "low-compute" interactions itself and forward the relatively infrequent major computations to the large machine. (p. 204)

In particular, Licklider noted the need for a significant amount of computer power at the user's location for the purpose of driving computer-controlled displays.

A second and related trend is the increasing power and sophistication of data communication technology. In the mid 1960s it was possible to link terminals to computers by telephone lines and to perform batch data transfers between computers. The late 1960s and early 1970s saw the advent of specialized, digital communication networks, and increasing sophistication in the kinds of communication that could take place over such networks. At present, there are long distance packet-switching networks such as the ARPAnet, local high bandwidth networks, networks that depend on the special characteristics of satellite communications, and short-range radio-based networks such as the packet radio systems mentioned in Section 2.1.

These networks can now do much more than support terminal linkage and simple batch data transfers. Sophisticated protocols have been developed that permit a rich variety of communications to take place between machines, between people, and between people and machines. At the same time, the decreasing cost of digital hardware has led to a proliferation in the number and types of computing elements attached to digital communication networks. What is emerging is a new computing/communication milieu involving many users and many machines, all linked by sophisticated digital communication networks (Nickerson, 1980).

These distributed, network-based systems follow a number of different architectural approaches, differing substantially in key properties. Thus, where there was once but a single, centralized architectural approach toward interactive system design, there is now a considerable variety of approaches.

Given the current and anticipated trends in the continued miniaturization of both processors and computer memories, it should be possible in the near future to package a very large amount of computing power and local storage within a portable terminal. The question arises as to what kinds of computational and storage needs should be satisfied locally and what kinds should be obtained from shared resources when the answer to the question is not dictated by financial considerations.

System architecture has human factors implications because the way in which the parts of a system are distributed will affect the system's properties as perceived by a user. Although this is generally understood, research that focuses on the human factors aspects of system architecture has lagged behind research on the various supporting technologies. Hence, the importance of this area in military research programs.

Among the human factors implications of system architecture are the following:

- Along with the emergence of distributed systems have come new applications. Message and teleconferencing systems, in particular, have considerable importance to the military as an essential element in command and control. Although such systems can be built on a centralized model, their ultimate power can only be realized by a distributed approach. In Section 2.1, we explored some of the human factors problems of message and teleconferencing systems.

- Distributed systems have a high potential for survivability in battle and other adverse conditions. The technology that permits such systems to survive has been studied in considerable detail, but the human factors of a system under survival conditions have not.

In Section 5.2, we consider some of the human factors aspects of survivable systems.

- o Responsiveness is an important system characteristic from a human factors point of view, and the way the parts of a system are distributed will have a strong impact on its response to user requests for service. In Section 5.5, we discuss some of the issues relating to system responsiveness.

- o One of the implications of the decreasing costs of microcomputers is the economic feasibility of developing machines with architectures that are qualitatively different from that of the serial processor design that has dominated the industry since its inception. Machines with a large number of processors working in parallel can now be built economically. However, effective use of these machines will require the development of new ways of thinking about problems. Historically, the conventional way of approaching a complex problem has been to break it down into subproblems that could be solved, or steps that could be executed, in sequence. Relatively little thought has been given to the question of how to break complex problems into parts that can be worked on simultaneously. We believe that the increasing availability of parallel processing machines will provide the motivation to develop new approaches to problem solving based on parallel processing algorithms. The development of such algorithms will require that problems be thought about in qualitatively different ways than they have been thought about in the past. The study of this "paradigm shift" is another opportunity for basic research that relates very directly to information technology and its future utilization.

5.2 Survivability

A powerful property of distributed systems is their potential survivability under adverse circumstances. For example, most packet switched communication networks can survive the loss of one or more nodes or branches, often without any loss of information en route at the time of outage. This ability follows from two fundamental properties of the design of these networks:

- o Most such nets possess sufficient redundancy so that an alternative route can be found around a failed node or branch.
- o The nodes of such a net (actually small computers) possess sufficient intelligence to exchange information on the network's condition, detect failures, and cooperate in devising alternate routes around such failures. Each node has at all times an accurate working picture of the state of the network and is capable of modifying its operation to adjust for changes in that state.

At a higher level, computing systems built from distributed elements have a similar potential for failure tolerance, and this potential arises from the same two fundamental design principles. Sufficient redundancy is provided in processing and information storage elements so that the system can survive the loss of one or more elements of each type. The redundancy is taken advantage of through information exchange among the system's active elements, which, working together, detect failures and alter the systems operation accordingly. In a military setting, these design principles can lead to exceptionally robust systems. With sufficient redundancy, a distributed system can remain operational through the loss of many of its parts.

An interesting situation arises when the damage sustained by a distributed system partitions that system into two or more components, each able to function alone, but totally cut off from the others. Research is underway to devise methods for the reconstitution of such decoupled networks.

Packet radio networks (see Section 2.1) present an even more dramatic situation. Since the nodes of such a network may be in motion, the network's topology can be expected to change over time. In addition, nodes may enter or leave such a network, and nodes within the net may be put out of operation in a battle situation.

Progress has been made toward ensuring the survivability of distributed systems and intensive research continues on this subject. However, so far as we know, little attention has been paid to the human factors of a distributed system under survival conditions. The result is that we may be designing systems which survive but fail to remain usable in crisis.

In active military service, a distributed system will be likely to support command, control, and communication functions that include communication between users, access to data, and the exercise of processing capabilities. Under conditions of partial failure, some portion of these capabilities will be lost. As we see it, there are two key sets of human factors concerns in such a partially functioning system. One relates to the matter of maintaining orderly communication among those parts of a military unit that remain connected by the system. The second is the problem of maintaining system operation in the face of lost data or processing capabilities.

One of the greatest risks in a crisis situation is the danger of confusion. A military unit may sustain serious damage but nonetheless carry out a successful mission provided the surviving parts can continue to function in orderly, coordinated fashion. Such coordinated operation in turn depends on continued communication among the surviving parts of the unit, and, equally importantly, on what might be termed a collective self awareness. That is, the unit's key members must have a clear mental picture of what parts are still active, where they are located, what action they are taking, and so on. Furthermore, this picture must be kept up to date in continuous fashion as the engagement proceeds.

Such collective self awareness can be obtained by manual polling of a unit's membership, but this consumes valuable time, and there is a potentially much better way to achieve the same result. The key lies in the basic survival mechanism of the system itself, which depends on a self awareness analogous to that needed by the system's users. As indicated previously, the active elements of a distributed system exchange information from which each element derives a continuously updated picture of the system's current state. This same mechanism, already built into the system, could be used to generate a state picture suitable for the human user.

Research on ways to provide military units with such a state picture is undoubtedly needed. Specific research questions that should be considered include the following:

- o The state picture generated by the system itself is limited to a simple indication of the status (active/inactive) of each element. Such a picture alone would help the user by providing an indication of what parts of his unit could currently be reached through the communication system. However, it seems clear that a more valuable picture could be obtained by entering additional information into the system and allowing it to be propagated among the system's active elements. What additional information would be useful to the system's users? Physical location, movement (if any), and intention suggest themselves as likely candidates.

- o How can such additional information be obtained by the system? Should the user enter it into his terminal, or can some parts of it (e.g., location, motion) be discovered by the system through communication among the terminal nodes? How can the terminal be designed to facilitate entry of the appropriate information?

- o How can the state picture best be presented to the user? A map format suggests itself in the case of display terminals, but what about the case of terminals not equipped with displays?

The second set of human factors issues relating to partially functioning systems concerns the partial loss of data and processing functions that may be suffered when a distributed system sustains damage. This is a relatively new area for human factors research, since the centralized systems that have predominated up to now are likely to be either fully operational or completely shut down. The problem, as we see it, is to anticipate the types of failures that may occur and provide for them in the design of the system. The art of building failure-tolerant systems of any kind depends on this basic approach. The following are some of the human factors questions that arise.

- What is the best way to present a situation (or other data) when some of the supporting data base is lost or no longer timely? This breaks down into two sub-questions -- how to design robust presentation software that can operate in the face of a loss of data (a software engineering question) and how to design display formats that take advantage of valid data while making clear what's missing or less valid.

- Can trend display notions be employed to extrapolate forward from the last known condition of some part of the total situation? If so, then a key human factors concern is to make clear the uncertain nature of such extrapolated information as contrasted with the more valid, up-to-date data available from those parts of the system that remain operational.

- What can be done to adapt to the loss of active system functions?

An interesting parallel exists here between the military situation and the techniques that are applied in air traffic control. The devices employed to monitor and control air traffic have evolved from manual plotting boards and flight strips (simple plastic blocks representing each flight) to primary radar, to transponder-assisted radar, to radar assisted by altitude reporting transponders plus computer tracking and aircraft data display. The basic approach is to retain all historic "layers" of the evolving technology in more or less active status so that as one fails the traffic controllers can fall back upon a previous layer. As it happens, each successive layer of advancing technology appears to be somewhat more frail than the one before, so that under failure conditions what one usually encounters is a reasonably orderly retreat back through the evolution.

In order for this approach to work, of course, it is necessary that the controllers be trained to work with the system at all its levels of technical sophistication, and, further, that they be practiced in rapidly shifting from one level to the one before. Recent failures of the topmost, computer-assisted layer have caused great concern among controllers, suggesting that the level of preparedness may not have been sufficient to permit the controllers to gracefully assume a fall-back position. Thus, although it appears possible that a similar approach could be employed in the military, it also seems clear that training is critical.

Extrapolating from air traffic control to the military, the foregoing discussion suggests that two criteria must be met. First, the design of interactive computer systems for the military should anticipate the possibility of "layered degradation." Allowance should be made for both partial and total loss of system capabilities. Manual systems and/or more primitive automated systems should be available in the event of total failure of the primary system.

Second, the key human factors concern under conditions of degraded performance is that the human users should be fully prepared to fall back to the more primitive support systems that will be called into play. This means careful training under conditions that simulate to the extent possible all failure conditions that may occur.

5.3 Interoperability, Inter-system Compatibility

There is currently great concern for interoperability among military systems and components at the hardware, software and data communication levels. The same concern does not, however, appear to extend to the level of the human operator of an interactive system.

In their survey of Battlefield Automation Systems, Synectics Corporation (1980) examined five automated systems for Army field use with particular reference to the human interface characteristics of those systems. Their report focuses on inter-system differences that can reasonably be regarded as unnecessary, that is, differences that do not arise from the fundamental differences in purpose or function among the systems. The report shows an array of unnecessary differences in categories such as the following:

- o Command types
- o Command Entry Method
- o Help/User Aids
- o Execution of transactions
- o Use of nonalphabetic symbols
- o Use of Boolean, relational, or logical operators
- o Terminal keyboards
- o Design and layout of display screen

Not only do the differences show up among systems, but <u>within</u> systems as well. An especially dramatic example concerns function keypads for use in one system (TACFIRE) at the division and battalion levels. Although 47 of the 64 possible codes are common to both keypads, the report points out that "Only 19 of these are in the same location; 28 are in different locations." As the report indicates, in something of an understatement, "Any user familiar with one menu, however, would become confused when trying to use the other."

The negative effects of such differences include unnecessary training for operators of more than one system, operator inefficiency, and potentially damaging errors. One way to summarize the problem is to observe that mental capacity that should be available for concentration on the functions to be performed is pre-empted by the need to remember unnecessary differences in the human interface.

It seems clear that some research should be devoted to the problem of system interoperability at the human level. A useful framework for attacking the problem may be derived from the technology of data communications. From this point of view, the human operator can be viewed as one node in a communication network in which the other nodes are the dissimilar systems with which he must cope.

The designer of a data communication system faces the problem of establishing orderly communication among potentially dissimilar nodes or subsystems. The basic technique is to design protocols, or rules, for the orderly exchange of information. If two subsystems can be designed to adhere to the same protocol, then they can communicate, no matter how different they may be in other respects. It is this basic technique, for example, that enables computers built by different manufacturers to exchange data through a digital network.

From the point of view of this communication system model, the present chaos in human interface design can be thought of as a world lacking in standardized protocols. The human user is asked to interface with a variety of other systems, each obeying a different communication protocol. This means that he must not only learn the various protocols sufficiently well to become adept at their execution, but also keep them sufficiently separate in his mind to avoid confusing one with another while operating a system. What is likely to make this all the more difficult is that many of the differences between protocols are apt to be subtle, hard-to-remember ones that are nonetheless critical for successful system operation. This situation will place a considerable processing burden on any system, human or otherwise. The remedy is to standardize on the protocols for communication between users and machines, just as is done with communication between machine and machine.

Beyond this, communication technology may have further insights to offer for the user-computer interface. As the art of protocol design has advanced, a second fundamental notion has emerged, and that is the concept of <u>layered</u> protocols. The idea of layering arises from the internal complexity of the communication that takes place between two systems, and from the desire to standardize to the extent possible even though multiple purposes may need to be served by inter-machine communication.

At a basic level, protocols will be needed to cope with the electrical properties of the communication link that binds the systems together. Higher level protocols will be concerned with error control, buffering, and other problems of digital information transfer across the link. The highest level of protocols will make use of the levels below to support the applications for which communication was needed in the first place.

It is at these highest protocol levels that differences are most likely to show up, reflecting fundamental differences in the purposes served. For example, a protocol for file transfer between two machines may differ from one that is intended to support message transport, or from one the goal of which is to support real-time teleconferencing. However, all three of these application-level protocols will rest on a common base of link-level and information-transfer protocols.

An essential property of layered protocols is the independence between layers. A given layer functions without concern for the inner workings of the layers below, or how it is being used by the layers above. It is this fundamental property of protocol design that makes it possible for many different applications to share the same set of lower-level protocols.

There are several reasons for believing that these protocol concepts may be usefully applied to the human interface of an interactive system. First, the interface does indeed appear to have layered, hierarchical properties. To the extent that this is true, the protocol approach will provide a useful descriptive tool, capable of clarifying the thinking of interface designers and analysts. Second, the independence between protocol layers means that the inner workings of a layer can be changed at will, provided its external appearance remains constant to the layers above and below. Finally, and most importantly, the protocol approach may help clarify what aspects of the human interface can be held constant across systems and what aspects must differ due to basic functional differences among systems. The following paragraphs represent a first attempt at partitioning the human interface into a set of independent layers.

Layer 1 consists of the physical interface, the artifacts of user-machine communication: keyboard, function keys, display screen, light pen, touch panel, printer, etc. Standardizing on this level of physical equipment would provide a foundation for protocol standards at higher levels. What suggests itself is a standardized, modular terminal design, modular in the sense that not all features (e.g., printer, touch panel) need be present in all units, standardized in the sense that if present, a feature adheres to a standard design and location in the total ensemble.

Research would be required to develop effective standard designs and to answer the question "how standard is standard enough?" For example, if function keys are present, must there always be precisely the same number of keys in the same location and layout, or can one or more of these parameters be varied within bounds? The goal of this research would be to develop a terminal design that would be comfortable and efficient in any of its variations and that would enable a user to shift from variation to variation without need for retraining.

Layer 2 consists of the language syntax that governs the transfer of information across the physical interface of Layer 1. Concerns at this level focus on basic choices of command input: commands typed on the keyboard, entered by function key, selected by touch panel from a menu, etc. Within each basic command entry mode, it would be necessary to consider numerous details. For example, with keyboard entry, how are commands terminated? Can a command line be edited prior to entry, and if so, what are the editing conventions? Is character-by-character interaction provided such that partially typed commands can be extended by the system? What confirmation is required for potentially dangerous commands? Are help features provided within a command line to cue the user? How are arguments indicated? Is there a subcommand mode? At this level it would also be appropriate to consider standard formats and assignments for display windows, command menus, and so on.

Again, the goal is to permit comfortable user transitions from one system to another, and one way to promote that goal is to standardize on as many of these command language features as possible. To the extent that any of these features differ among systems, the user is presented with a learning problem made all the more difficult because it may represent one of a few minor but pernicious differences between otherwise similar interface designs. As with the physical level, it is not clear how rigid these standards must be. It is a research question to determine, for example, whether keyboard, function key, and menu selection should be provided as equally acceptable options. What does seem clear, however, is that if one of these options is selected, it should be held constant across system boundaries.

Layer 3 relates to semantics. It describes the environment perceived by the user to exist within the computer system -- objects, their structure and properties, and the basic operations that can be carried out upon objects. The goal here is to standardize on those aspects of system design that can reasonably be held constant across systems. Objects such as files, records, fields, messages, matrices, and indices should have identical names, identical properties, and identical forms for printing or display to the user. The operations for creating, deleting, moving, and modifying such objects should also have identical names and semantics.

Again, it would not necessarily be possible or desirable to have all systems be precisely identical in all respects. It does seem desirable, however, for the content of such systems -- operations and objects -- to be drawn from a common pool. Research is needed in this area to determine what objects and operations are needed and what their properties should be.

Layer 4 is concerned with the application to which the system is to be put. It relates to the actions that can be taken in furthering that application. To the extent possible, each such action represents a composition of operations carried out on objects at the layer below.

With the subsidiary layers in place, design at this level should follow in quite straightforward fashion, guided by the application in hand. An interesting question is the extent to which different systems will in fact differ, once appropriate standardization has been achieved at the lower levels of user-machine protocol. It is the suspicion of the authors, based on experience with other interactive systems, that those residual differences truly derived from differences in application will not be particularly great.

Clearly this scheme needs to be developed in much more detail, but it appears to have the following useful properties.

- It isolates into one layer (layer 4) the essential differences between systems. To a substantial extent all layers below this could be standardized across a considerable variety of systems.

- To a considerable extent, experimentation can be carried out within some of the layers without affecting the others. For example, substantially different methods of command input (keyboard entry, function keys, menu selection, etc.) can be employed without affecting the layers above.

- The system environment -- a critically important layer, since it governs the user's model of the system and its operation -- is isolated from the others. In our experience, this portion of the human interface has received less than sufficient attention.

5.4 Security

As computers have found increasingly widespread use in the military, it has become necessary to consider security issues as a factor in system design. Depending on the approach taken, security may have a negligible or a profound impact on the human factors of an interactive system. From a human factors point of view, the least troublesome current approach to security is to operate in a mode termed "system-high." System-high means two things. First, all users of a system that is operating in this fashion are assumed to have authorization to access all data in the system, from the dual points of view of security classification level and need to know. The entire facility is guarded against penetration from users outside of this limited group. Second, all information in the system is regarded as potentially having the highest classification of any information in the system. No attempt is made to segregate the information contained in terms of its security classification level or need-to-know designation.

Thus, although a system-high facility may be heavily guarded from the outside, once inside such a facility, the interactive computer system itself is much like any other. Information within such a system can be manipulated in any manner supported by the system's functionality. Proper adherence to security procedures is left up to the user, who can, for example, freely mix information of different classification levels if he deems that to be appropriate. Thus, from a human factors point of view, a system operated at system-high is affected by no more than the same issues and concerns as any other interactive system.

From a military point of view, system-high operation is less than ideal. It has the effect of locking up information at unnecessarily high security levels. One goal of military security is to ensure that users with a "need to know" have access to the data they need that are classified at or below their clearance level. Unless the user's clearance level matches the highest level of the system, he will be denied access to any information in the system, even though much of that information may be appropriate and desirable for him to have.

Thus there is a strong motivation to design systems that can not only contain data classified at multiple security levels, but also provide controlled access by users with differing levels of clearance. When this is done, the burden of security protection is shifted away from the surrounding enclosure and into the system itself.

As might be expected, such a shift of burden has human factors ramifications. To begin with, barriers must be built into the system to control information access in accordance with the user's clearance level. In particular, such barriers must prevent any user from gaining access to information classified above his clearance level. However, the main human factors concerns do not arise so much from the presence of these barriers as from a fundamental problem in computer science. The fact is that from a security point of view computer programs cannot be trusted to operate properly. This situation exists because there is currently no mathematically reliable way to prove that any but the very simplest program has been correctly written. Vigorous research is underway to correct this situation, but to date it remains a fact of life in secure system design. In a multi-level secure system, there exists the risk that a faulty program may cause information leakage from a highly classified level to a lower level within the system. Information "leaked" in this fashion could then pass out of the system and into the hands of users unauthorized to receive it. Such program faults may occur either through normal human error or deliberate sabotage on the part of a programmer. In either case, the effect is to jeopardize the integrity of the system.

At present, the only known approach to circumvent this problem is to place all data access in the system under direct control of a software "kernel" small enough to be reasonably verifiable as correct. In general, no other software in the system can be trusted. When put into effect in a working system, this solution places a heavy burden upon the user with a high clearance level who wishes to manipulate information stored in the system that is classified at several different levels. When examined in detail, manipulations of this type are found to contain many small elementary steps that could be regarded as potential security violations if initiated by the untrusted software, rather than by the user. To ensure that all is as it should be, the trusted system kernel must verify with the user that each such step was truly intended. The net effect is to clutter the human interface with what the user perceives as unnecessary and distracting "noise." The only alternative to such distraction is to restrict the manipulations available to the user, an approach that reduces the utility of the system.

Message handling highlights the problem quite effectively because of the structure of messages and the manipulations commonly carried out upon them. A single message may contain information at several security levels. It may be desirable to copy information from one message to another. It may be necessary to generate an unclassified reply to a classified message, and so forth.

Research continues on the technology of multilevel secure system, but to our knowledge relatively little has been done with respect to the human interface for such systems. Given the growing importance of secure information systems, such research should have a fairly high priority within the military.

5.5 Responsiveness

The responsiveness of a computer system can have profound effects on its usefulness to and acceptance by a user community. One might argue that fast response is good, slow response is bad, and let it go at that. What this argument overlooks is that rapid response time has a cost, and, therefore, tradeoffs are necessary in setting the desired response level of a system. Furthermore, there may be good reasons of providing for levels of responsiveness that differ depending on the function performed.

All of this is understood in an intuitive and rudimentary way by system designers, but so far as we know, there have been few studies that attack the problem in systematic fashion. Such studies might result in guidelines that provide the designer with a more concrete and useful foundation than intuition. In the following paragraphs we discuss aspects of responsiveness that might be considered in research studies.

5.5.1 Basic Capacity

A mode of system failure that is often experienced in crisis situations is saturation of system capacity to the point where a system is brought to a standstill and becomes totally useless. Such saturation conditions were experienced in the domestic telephone network immediately after the assassination of President Kennedy in 1963, and during the Northeast Power blackout in 1965.

Commercial communication networks are designed with enough capacity to handle load conditions above normal out to some statistically reasonable level, but not under extreme crisis. The telephone network saturated and stalled in the two crises just mentioned because it was not designed to handle the loads presented to it. This is regarded as normal and acceptable with commercial systems, and in fact very little harm probably came of these two breakdowns.

Experience with military communication systems suggests that they exhibit the same behavior under extreme loads and for the same reasons. However, here the behavior is not acceptable. In fact, since the basic goal of a military unit is to handle crises, military systems that saturate under crisis conditions can be said to be inadequately designed.

The solution is an engineering one, not a human factors one. Such systems must be designed with enough capacity to handle crisis loads. Human factors concerns arise from the fact that user behavior may be effected long before a system becomes so loaded as to actually saturate. Studies are needed to determine user and system performance under a wide range of load conditions.

Of course, capacity has a cost, and systems designed for high capacity are bound to cost more than other systems. It will be up to the military planners and ultimately up to the country as a whole to decide how much to invest in military communication and information systems. However, no matter what decision is finally made, a careful analysis of needed capacity and studies of effects on performance of having something less than that will at least serve to make clear the consequences of that decision.

5.5.2 Responsiveness and the Psychology of Acceptance

Users find responsive systems more pleasant to use than systems that exhibit poor response times. Below some level of performance, users will begin to avoid the use of a system. (See Nickerson, 1969, in press b, for discussions of some of the factors relating to effects of response time on user attitudes and behavior.) At still lower levels of performance, one may encounter outright rebellion in the user community. This description characterizes the behavior of voluntary system users, but user satisfaction is bound to have an impact even in military contexts where system usage may not be voluntary. At the very least, one would expect a negative effect on morale when usage of an unresponsive system was mandatory.

In recent experiments at Camp Smith, Hawaii, these expectations were largely borne out by unexpected response-time problems in a message system under trial use. Until its performance could be improved by an infusion of new hardware, the system went largely unused, and the mandatory experimental sessions were viewed with considerable distaste.

The point of these observations is that responsiveness has psychological implications that may damage the utility of a system beyond any direct impact on time to get a job done. Given that fast response generally comes at a cost, it might be useful to study the psychological reactions of users in more detail than a simple observation that poor response is looked at with disfavor. The following paragraphs offer some points of departure for such research.

There is some reason to believe that users have rational expectations with regard to responsiveness, i.e., that the level of responsiveness that will be tolerated by a user depends on his perception of how much work the system has been asked to perform. In our own experience, users will tolerate considerable delay for completion of a long file search, but expect very fast response on a simple text editing operation such as inserting a character or word into a body or text. When simply typing in the characters of a command, a full-duplex system had better echo those characters quickly or there is apt to be strong user dissatisfaction.

Of course, the work believed necessary by the user may differ substantially from what the system is actually required to do. For example, depending on the software design and how much text had already been entered, the simple edit just suggested might require substantial shuffling of information in computer memory. A useful investigation would ccrrelate user response tolerance with, on the one hand, the work believed necessary and, on the other, the work actually necessary to accomplish a function. The results of such an investigation could provide the system designer with a predictive tool for gauging where effort should be put to achieve fast response.

Even if substantial processing delays are necessary and accepted as such by the user, there is some evidence to suggest that the typical user needs reassurance that the system is, in fact, carrying out work on his behalf. When a terminal sits inert, with no sign of inner activity, there tends to be a thin edge of anxiety that the system may have ceased operation, or worse still, abandoned one's job in favor of others. One technique that has been found useful in these circumstances is for the system to post short messages at frequent intervals reassuring the user that processing is in fact underway. If such messages can also indicate the proportion of work accomplished up to that time, then the user receives even more solid confirmation of where he stands.

5.5.3 Architecture

System architecture can play a strong role in shaping the response patterns of an interactive system. If the designer has the freedom to manipulate the basic architecture of a system, then he may be able to guarantee that a system will have acceptable response times. An investigation into the human factors of system response should certainly include a careful examination of the impacts and opportunities that arise from the system's architecture.

A difficult problem arises with the centralized, time-shared systems that have predominated up to now. Almost all time-shared operating systems have some form of control mechanism that schedules the allocation of key system resources to user jobs. Such resources include the central processing unit, memory, and input-output channels. There are many possible algorithms for scheduling resources to jobs, but almost all of them seek to make sure that all available resources are meted out and none are left unnecessarily idle.

When such systems are subject to fluctuating demand, as is usually the case, the individual user will receive varying portions of those key resources. The net effect is a fluctuating system responsiveness that closely tracks the total load placed on the system.

One widely held view is that such shifting responsiveness is disturbing to the user. An approach that has sometimes been suggested as an antidote to this problem is to design scheduling algorithms so as to stabilize a system's perceived responsiveness, thus minimizing the impact of fluctuating load conditions. This would seem to make sense only if response time could be stabilized at a satisfactory level.

A countervailing view suggests that once the user understands the inevitability of load fluctuations on a centralized system, he is apt to accept this phenomenon as a fact of life. Better than average response will please such a user, especially if he has made an effort to access the system at times when loads are typically light. Average load conditions may prove acceptable; heavy loads will be likely to cause resentment. This view suggests that if the maximum allowable load is controlled at a satisfactory level, then load fluctuations below that maximum will cause little trouble. The validity of these differing views could be investigated by some relatively straightforward experiments.

As computer hardware comes down in price, it will be increasingly possible simply to buy one's way out of load problems by acquiring sufficient hardware to satisfy the heaviest loads expected. This is true regardless of the architectural approach that governs a system's design. However, distributed architectural approaches offer some added response control possibilities that make them especially attractive. Assuming that users do indeed have "rational expectations" with regard to response time, then this phenomenon can be taken advantage of by placing certain functions that require fast response in the terminal hardware that is dedicated to a single user. This would include command language interpretation (at least at the syntactic level), text editing, display generation, and other functions that require the kind of processing support that can be provided locally. Functions requiring more intensive processing, such as file searches or extensive text formatting operations, can be relegated to a larger, mainframe processor shared by many users.

This approach leaves the user with a dedicated terminal having response characteristics that are both rapid and stable. The system always appears "alive" because the terminal always responds to user input. This approach appears especially well suited to the current generation of intelligent terminals, which have sufficient processing power to handle the light-duty interactive functions, but not the heavier process- or storage-intensive operations.

Multiprocessing techniques provide another powerful design tool that can be used with either centralized or distributed systems. Multiprocessing makes it possible to carry out simultaneously several different tasks on behalf of a single user. For example, a user request for an extensive file search triggers the creation of a second process that carries out the search while the original process continues to accept user command input. If the system is a centralized one, then the two processes will share that single processing resource. If it is a distributed system, then the file search process is likely to be carried out on a processor separate from the user's terminal. In either case the original process continues to devote its attention to direct interaction with the user.

The situation is not unlike that of an executive with a chief assistant and one or more subordinate assistants. Through the mediation of the chief assistant, tasks are delegated to subordinates. In the meantime, dialogue continues with the chief assistant. As tasks are completed, the chief reports this to the executive, who may modify his subsequent behavior accordingly.

The mapping of functions onto distributed systems, and the use of multiprocessing techniques are under investigation from an engineering point of view, but so far as we know, little study has been carried out with a specifically human factors orientation. Some interesting questions suggest themselves in considering these techniques:

- Can the primary user-interaction process directly support sufficient functionality to satisfy the user that he has the attention of a useful assistant? If all functions are dedicated to a centralized machine or subsidiary processes, then the interaction process will offer little more capability than an unintelligent terminal. On the other hand, if too much functional burden is placed on this process, then the advantages of distributed functionality will have been lost.

- With multiple processes simultaneously carrying out different assignments, the user is left in a position somewhat analogous to that of a juggler. Under these conditions the dynamics of an interactive session will be substantially different from the linear stream of tasks that is currently the norm. The efficiency advantages of such multi-tasking seem clear, but does it pose problems from a human factors point of view?

6. METHODOLOGICAL ISSUES

6.1 Major Approaches to the Study of User-computer Interaction

Studies of user-computer interaction that have been done to date can be grouped, at least roughly, in terms of which of the three following methods have been employed: (1) observation, (2) controlled experimentation, and (3) modelling and simulation.

6.1.1 Observation

Here the approach is to take measurements and gather statistics on real systems with real users working on real problems. An advantage of the approach is the fact that objective measures often can be made unobtrusively by the system itself. This is important inasmuch as interactive systems tend to be costly, and typically the motivation is high to keep them productively busy. A disadvantage is the fact that one has little or no control over the parameters whose effect one may wish to study.

Some investigators have looked for behavioral invariants and have attempted to find functional relationships that would characterize user-computer interaction in general ways. The number of observational measurements that have been made by investigators of user-computer interaction is large. They include the number of commands executed per unit time, interaction cycle time, task turnaround time, output/computer-time ratio, user and system response times, work session duration, console-time/CPU-time ratio, statement interpretation rate, overhead rate, number of user input lines per unit time, and rate of user requests (Bryan, 1967; Carbonell, Elkind & Nickerson, 1968; Grignetti & Miller, 1970; Grignetti, Miller, Nickerson & Pew, 1971; Scherr, 1965). Perhaps the most reliable finding to date is the tremendous amount of user and system variability. It is not uncommon to find individual differences on a given performance measure in ratios of 10:1 or 20:1, even when individuals are working in, presumably, identical situations.

6.1.2 Controlled Experimentation

A few investigators have attempted to run controlled experiments on user-computer interaction. This approach has the advantage that one can manipulate system parameters and study the effects in a systematic fashion. It also has its problems, however. Perhaps foremost among them is the fact that the systems of interest tend to be large and costly, which means that unless the experimenter is in an unusual situation, he may find it difficult to have access to such a system for experimental use. Moreover, often because of the fact that computer systems are so costly and the developers are vulnerable, there may be strong vested interest in obtaining a particular result. Finally, even if these sorts of problems did not exist, there is the problem of the generality of one's findings. Too often, results from experiments cannot be generalized beyond the particular systems, users, and situations with which they were obtained. In studying the relative merits of different programming languages, for example, one must take care that the results are not contingent upon the particular programming problems that were selected or on the specific subjects who wrote the programs. To get around the problem of programmers bringing different types of programming experience to the task, some investigators have used subjects who lack any programming experience, but this approach has its limitations also. What proves to be best for novices will not necessarily prove to be best for experienced programmers or users. Similarly, what proves to be most advantageous with respect to one programming problem may not be so with respect to other problems.

6.1.3 Modelling and Simulation

Performance evaluation has always been recognized as an important component of the system development cycle. What has been less generally recognized is the importance of performance prediction. What one would really like to be able to do is to predict in advance of system implementation the performance of the equipment, the user, and the user-machine complex. Further, one would like to be able to predict how that performance would depend both on the characteristics of the system and on the situation in which it is used. One would especially like to be able to predict performance in high-demand, stressful, crisis situations.

This is, of course, a very difficult problem. In order to predict performance with any degree of precision, one must have a good model of the system. Moreover, the only way to validate a model is to compare its predictions against results obtained when the system that is being modelled is actually exercised. The value of modelling, however, derives in large measure from the fact that on the basis of modelled results one can make decisions that effect the future development of the system. This being the case, many of the predictions of the model are inherently untestable, inasmuch as some of the versions of the systems that are modelled are never produced.

Notwithstanding such problems, the modelling approach has much to recommend it. Moreover, validation is not an impossible problem. Although not all predictions of a model can be checked, some of them can, and like any other theory, the credibility of the model increases with the amount of evidence that can be presented in its favor. The more the model proves to make accurate predictions in those cases that are testable, the more one is likely to have some confidence in its probable accuracy in those cases in which it is not testable.

A good example of the use of simulation to study human performance in command and control systems is provided by some of the work of the Systems Effectiveness Branch of the Aerospace Medical Divisions Human Engineering Division (Mills, Bachert, & Aume, 1975). They have developed a simulation facility with five user-computer display consoles, and have used it to study the human factors design considerations of the BUIC System, and the AWACS early warning system. They have implemented an RPV command and control system having four en-route controller positions and a terminal control position. Many system variables have been studied including the number of vehicles per operator, the RPV flight-control algorithm, and the frequency of actual path up-date. In each case, user-computer system performance data have been collected on experienced teams of controllers, thus providing important information for future design trade-off studies.

Several developmental threads over the past 15 years are beginning to come together to provide an ability to represent integrated human performance analytically, either in mathematical or computer form. Describing-function and optimal-control models of manual control have proven their usefulness as design aids. The work of Siegel and Wolf (1969) has led to the SAINT Simulation language specifically designed to represent human task performance in event-oriented simulations. The Human Operator Simulator (HOS) has been used to attempt to integrate elementary models of human performance into larger task-related simulations. The state of model development for large-scale multi-person systems remains crude. It requires detailed understanding of the sequence and timing of specific procedures together with methods for assigning priority structures to the sequence of activities undertaken by the operators.

Besides construction and validation of models applicable to specific task contexts, there is still much methodological work to be done. As the scope and complexity of models increase, it becomes much more difficult to understand what it means to validate a model. So many parameters are involved that it is no longer possible to test experimentally even a representative sampling of the range of conditions that might be explored. We must redefine what we will accept as validation if we are not to reject the modelling concept altogether for large scale systems application.

Models can be developed by starting with the highest level system goals and working down in the level of detail until the model is predicting results at the desired level, a so-called top-down approach. Alternatively, one can begin from first principles of human information processing and control, and build up to the desired level of task integration; which is a bottom-up approach. Identifying the conditions under which each approach is applicable is a problem for further research.

Finally, in the current state-of-the-art, each new context requires the generation of a new model practically from ground zero. Research is required to identify the general structure of system models applicable to classes of tasks that can then be particularized for specific applications.

Understanding human behavior in monitoring and supervisory control presents its own special problems. Human factors specialists have a variety of resources in the form of limited theories and methods that should be brought to bear on this class of problem, but a great deal remains to be done to apply them in this context.

In the area of real-time monitoring and control of continuous dynamic processes, the optimal control model (Baron & Kleinman, 1969) has been shown to describe suitably the perceptual-motor behavior of closed-loop systems having relatively short time constants. Limited experimentation suggests that this class of model may be broadened to represent monitoring and discrete decision behavior in dynamic systems in which control is infrequent (Levison & Tanner, 1971). Some attempts have also been made to extend this work and to explore its applicability to more complex systems.

In the field of vigilance research, which is concerned with human behavior in systems in which the detection of weak and infrequent signals is required, much is known about what parameters of signal presentation affect performance, and the signal detection model (Green and Swets, 1966) has been shown to be useful in analyzing behavior in such cases. But, again, its applicability has not been evaluated in tasks in which "signals" are represented by complex patterns of activity.

Typically, the methodology for understanding the task and information requirements for complex tasks has employed task analysis, activity analysis, and the development of operational sequence diagrams (Meister & Rabideau, 1965). Recent developments have suggested that these procedures can be augmented by the collection of directed verbal protocols in order to understand better the cognitive model employed by the operator for drawing inferences from diverse sources of data (Stevens & Collins, 1977).

In Denmark, Rasmussen (1976) has suggested that the appropriate way to represent the cognitive operations underlying human supervision of process control operations is through semantic nets and concepts derived from computer simulation and artificial intelligence models. This appears to be a very promising approach, but has not as yet been pursued in the United States.

6.2 Some Other Methodological Issues

6.2.1 Cognitive Workload Measurement

A classic problem in the area of human performance is that of measuring workload in various situations. This is an especially challenging problem when the workload is primarily or exclusively cognitive. Workload measurement techniques have been developed and applied to a variety of task situations, although to our knowledge none has been applied to the problem of measuring workload of the user of a computer system. The general literature on divided attention should facilitate the development of appropriate models and measurement techniques.

6.2.2 On Identifying What is Wrong with Existing Systems

Sheil (in press) has made the point that there are few tasks less demanding than that of lampooning the interface designs of existing computer systems. It is one thing to point out obvious defects, however, and quite another to design a system that has none, especially given the economic and other constraints within which designers frequently work. If one is inclined to play the critic role without offering some useful guidelines for designers along with one's criticisms, one might do well to reflect on Sheil's reaction to this kind of enterprise: "It is downright dishonest to suggest, on the basis of several such examples [of systems with poorly designed interfaces], that the difficulty of using computer systems reflects simple incompetence or insensitivity on the part of their designers and that these problems would simply not arise were those designers as intelligent (sophisticated, learned in human factors, etc.) as 'we' are .... The fundamental issue is that criticisms based on an appeal to absurdity are entirely unprincipled. They appeal to one's 'common sense' (and arrogance and snobbery) without giving the slightest useful indication of what a designer is supposed to do in response. They are both sloppy science and a maximally ineffective way to conduct a dialogue with a technical community." Furthermore, Sheil argues, the advice that designers need is something more specific than vague generalizations of the sort that promote such features as uniformity and simplicity as design objectives without indicating how those objectives are to be attained.

We believe Sheil's point is well taken. It speaks to the issue of how human factors researchers should relate to system builders and, in particular, to the importance of presenting opinions and findings in an appropriately modest and objective manner. This is not to argue that critiques of existing systems are not useful or desirable, but simply to note the importance of bearing in mind that saying what is wrong with a system usually is a far easier task than developing a working system in the first place, and easier also than saying exactly how a system should be changed in order to correct the deficiencies that are identified, without creating more egregious ones in the process.

6.2.3 Software Evaluation

Software evaluation is one key to effective integration of human factors requirements into computer software design. Identifying better designs depends on having methodologies for deciding which of a set of designs does a better job of taking into account human factors issues. Such a methodology could take one of several forms. It could be in the form of a design guide or checklist describing desirable software features. It could be the specification of an experimental evaluation procedure that is followed for each design to be considered. It could depend on having impartial observers or the users themselves complete a rating form that probes the ease of use. It could involve an activity analysis examining how a user's time is spent under alternative designs. In principle, these methods are not different from what might be applied to the evaluation of any system designed for human use.

One thing that is unique about software evaluation is that there are new dimensions to be specified that potentially affect the speed, accuracy, and acceptability of the resulting product. In particular, inasmuch as software products tend to represent the more logical and cognitive aspects of system performance, they tend also to require analysis and evaluation of the users' reactions in terms of more complex cognitive processes than conventional hardware systems.

While it is common to evaluate hardware systems by mocking up designs, working through procedures, or exercising a working simulation, these practices are much less common today with software evaluation. One reason may be that only recently have human factors specialists attempted to develop the tools to mock up software interfaces or to generate computer languages specifically adapted to simulating the user interface of a proposed design. The development of such languages should enhance the ease with which software evaluation can be accomplished.

It is perhaps the flexibility obtained by the introduction of software components into the overall system development process that makes the methodological problems difficult. It is difficult to define standard methods that are applicable across such a wide range of conditions of use.

6.2.4 Methodology for Computer System Design

It has been universally difficult to produce designs for the user-computer interface in interactive systems that meet human factors requirements for ease of use, effectiveness and minimum error. As Ramsey and Atwood (1980) have argued, it is not a solution to propose a human factors specialist on every design team, because there are many thousands of systems being designed and perhaps only a few hundred human factors specialists qualified to help design such systems. Some have proposed the development of design guides, but getting such guides used by systems analysts who are unaware that they have a problem is difficult. Alternatively, a design methodology may be specified that ensures that issues of design for human use get considered at each stage of the development cycle.

Such a methodology has many steps to be considered, from initial conception to final documentation. Key elements are methods for representing how the job is done now. The same methods may then be used to represent job performance, given the new design. Ideally, these methods would provide representation at the level of the task to be accomplished, at the level of the structure or model of the task as viewed by the human user, at the level of the functional system requirements, and at the level of the specific implementation features that will make the user's job as simple to learn and as easy to accomplish as possible.

It seems likely that an effective methodology will also include the development of the capability to simulate and test simulated versions of proposed hardware and software designs on real users before these designs are fully committed. It is relatively easy to develop such a simulation capability at the level of simple branching interactive dialogue. It is much more difficult to accomplish it when models of the system elements are required or where the ease of interaction is influenced by the structure or design of the data base or communications protocols.

Other methodological components are the evaluation tools that will allow competitive designs to be evaluated objectively with respect to human factors issues. From the point of view of military procurement regulations, the question is whether it is possible to specify in advance a set of human factors evaluation criteria that can be applied uniformly and objectively to assure that design proposals will be appropriately judged on the basis of probable effectiveness of system performance.

Finally, there is a need to bring these various components of a methodology together into a coherent package that makes the designer's job easier rather than adding yet another burden to the already complex design task.

7. REFERENCES

ASD, Modernization of the WWMCCS Information System (WIS). Prepared by the Assistant Secretary of Defense (Communications, Command, Control and Intelligence), 19 January 1981.

Baron, S., & Kleinman, D.L. The human as an optimal controller and information processor. NASA CR-1151, September 1968.

Bartee, T.C., Buneman, O.P., Gardner, K.A., & Marcus, M.J. Computer internetting: C3I data communications networks. Institute for Defense Analyses, Science and Technology Division, IDA Paper P-1402, April 1979.

Bavelas, A., Belden, T., Glenn, E., Orlansky, J., Schwartz, J., & Sinaiko, H. Teleconferencing: Summary of a Preliminary Research Project, IDA Study S-138, November 1963.

Brown, R.L. A content analysis of communications within army small-unit patrolling operations. Human Resources Office, Tech. Rep. 67-7, 1967.

Bryan, G.E. JOSS: 20,000 hours at a console -- A statistical summary. AFIPS Conference Proceedings, 1967, 31, 769-777.

Bylinsky, G. The Japanese chip challenge. Fortune, March 23, 1981, 115-122.

Carbonell, J.R., Elkind, J.I., & Nickerson, R. S. On the psychological importance of time in a time-sharing system. Human Factors, 1968, 10, 135-142.

CENTACS, Standardized Computer Resource Interface and Management Plan. Center for Tactical Computer Systems, U.S. Army Communications Research and Development Command, Ft. Monmouth, N.J., December, 1980.

Chapanis, A., Ochsman, R.B., Parrish, R.N., & Weeks, G.D. Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. Human Factors, 1972, 14, 487-509.

Chapanis, A., Parrish, R.N., Ochsman, R.B., & Weeks, G.D. Studies in interactive communication. II. The effects of four communication modes on the linguistic performance of teams during cooperative problem-solving. Human Factors, 1977, 19 (2), 101-126.

Chapanis, A. Computers and the common man. In Kasschau, R., Lachman, R., & Laughery, K. (Eds.), Psychology and information processing technology: A view of the future, in press.

Chaudhari, P., Giessen, B.C., and Turnbull, D. Metallic glasses. Scientific American, 1980, 242(4), 98-117.

Cuff, R. N., On casual users, International Journal of Man-Machine Studies, 1980 (12), pp. 163-187.

DARCOM, The Future of Electronic Information Handling at the FCC. Blueprint for the 80's. Washington, D.C.: Federal Communications Commission, October 31, 1980.

Dertouzos, M.L., Report on the Advisory Committee on Information Network Structure and Functions for the Executive Office of the President. Appendix F in the Future of Electronic Information Handling at the FCC, Blueprint for the 80's, Federal Communications Commission, October 31, 1980.

Forgie, J.W., Feehrer, C.E., & Weene, P.L. Voice conferencing technology program. Tech. Rep. No. ESD-TR-79-78. Lexington, MA: Lincoln Laboratory, Massachusetts Institute of Technology, March 1979.

Frankhuizen, J.L., & Vrins, T.G.M. Human factors studies with Viewdata. Paper presented at 9th International Symposium on Human Factors in Telecommunications, September 1980.

Glaseman, S., & Epstein, H. Design of a terminal subsystem providing user access to distributed data base management systems. Rand Corp., April 1978.

Green, D.M., & Swets, J.A. Signal detection theory and psychophysics. N.Y.: John Wiley & Sons, 1966.

Grignetti, M.C., & Miller, D.C. Modifying computer response characteristics to influence command choice. Proceedings Conference on Man-Computer Interaction, Publication No. 68, pp. 201-205, Institution of Electrical Engineers, London, 1970.

Grignetti, M.C., Miller, D.C., Nickerson, R.S., & Pew, R.W. Information processing models and computer aids for human performance: Task 2: Human-Computer Interaction models. Tech. Rep. No. AFOSR-TR-71-2845 (NTIS No. AD732913), 1971.

Johnson, R.C. Thirty-two bit microprocessors inherit mainframe features. Electronics, February 24, 1981, 138-141.

Jones, V.E. Final report of the software acquisition and development working group, prepared for the Assistant Secretary of Defense for Communications, Command, and Intelligence, July 1980.

Kahn, R.E. Submicron digital technology. Unpublished memorandum, Defense Advanced Research Projects Agency, Arlington, VA 22209, 6 June 1978.

Kasschau, R., Lachman, R., & Laughery, K. (Eds.), Psychology and information processing technology: A view of the future, in press.

Kelly, M.J., & Chapanis, A. Limited vocabulary natural language dialogue. International Journal of Man-Machine Studies, 1977, 9, 479-501.

Klemmer, E.T. (Ed.), Human Factors, 1973, 15 (5).

Leonard, M.G. Promise of RAM market pushes chip technology frontiers. High Technology, June 1980, 1(5), 57-63.

Levison, W.H., & Tanner, R.B. A control-theory model for human decision making. NASA CR-1953, December 1971.

Licklider, J.C.R. Man-computer symbiosis. IRE Transactions on Human Factors in Electronics, 1960, 1, 4-11.

Licklider, J.C.R. Man-computer interaction. In C.A. Cuadra (Ed.), Annual Review of Information Science and Technology, Vol. 3, Chicago: Benton, 1968.

Matiso, J. The superconducting computer. Scientific American, 1980, 242(5), 50-65.

Meister, D., & Rabideau, G.F. Human factors evaluation in system development. New York: Wiley, 1965.

Military Computer Family (MCF) Equipment, General Specification for U.S. Army Communications Research and Development Command, Ft. Monmouth, NJ, June 1980.

Miller, J.G. Living systems: Basic concepts. Behavioral Science, 1965, 10, 193-237.

Mills, R.G., Bachert, R.F., & Aume, N.M. Summary report of AMRL remotely piloted vehicle (RPV) system simulation study II: Results. Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, Ohio, AMRL-TR-75-13, 1975.

Monty, R.A., Geller, E.S., Savage, R.E., & Perlmuter, L.C. The freedom to choose is not always so choice. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 170-178.

Moran, T.P. Introduction to the Command Language Barrier (Tech. Rep. SSL-78-3). Palo Alto, CA: Xerox Palo Alto Research Center, 1978.

Myer, Theodore H., Future message system design: Lessons from the Hermes experience, Proceedings of Computer Conference Fall '80, IEEE Computer Society, September 1980.

Nickerson, R.S., Man-computer interaction: A challenge for human factors research. Ergonomics, 1969, 12, 501-517. Reprinted in IEEE Trans. Man-Machine Sys., 1969, MMS-10, 164-180.

Nickerson, R.S., On conversational interaction with computers. In S. Treu (Ed.), User-Oriented Design of Interactive Graphics Systems, Proceedings of ACM/SIGGRAPH Workshop, 14-15, October 1976, Pittsburgh, Pa. pp. 101-113. Also BBN Report No. 3499.

Nickerson, R.S., Adams, M. J., Pew, R. W., Swets, J. A., Fidell, S. A., Feehrer, C. E., Yntema, D. B., & Green, D. M. The C3-system user. Vol. 1: A review of research on human performance as it relates to the design and operation of command, control and communication systems. BBN Report No. 3459 (submitted to Defense Advanced Research Projects Agency, February 1977).

Nickerson, R.S., Some human factors implications of the blurring of the line between communication and computation. Paper given at the 9th International Symposium on Human Factors in Telecommunications, Red Bank, NJ, September 29-October 3, 1980. BBN Report No. 4577.

Nickerson, R.S., Fidell, S.A., Kalikow, D.N., Nuthmann, C.F., Feehrer, C.E., Selfridge, O.G., & Vittal, J.J. Feasibility analysis for a n w compu er system, BBN Rep. No. 4030. Cambridge, MA: Be ek and Newman, 1980.

Nickerson, R.S. Information technology in the 80's: A retrospective look at some looks at the future. In Kasschau, R., Lachman, R., & Laughery, K. (Eds.), Psychology and information processing technology: A view of the future, in press a.

Nickerson, R.S. Why interactive computer systems are sometimes not used by people who might benefit from them. International Journal of Man-Machine Studies, in press b.

Perlmuter, L.C., Scharff, K., Karsh, R., & Monty, R.A. Perceived control: A generalized state of motivation. Motivation and Emotion, 1980, 4, 35-45.

Pew, R.W., Sidner, C.L., & Vittal, J.J. MMI (Man-Machine Interface) design documentation: Representing the user's model of a system. Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, CA, October 1980.

Pew, R.W., Woods, W.A., Stevens, A.L., & Weene, P. Identifying information systems requirements for decision making. Report ESD-TR-78-169, prepared for Development Plans, Electronic Systems Division, Hanscom Air Force Base, MA, August 1978.

Phipps, C. Implementation of computer technology in the 1980s: A semiconductor perspective. In Kasschau, R., Lachman, R., & Laughery, K. (Eds.), Psychology and information processing technology: A view of the future, in press.

Ramsey, H.R., & Atwood, M.E. Man-computer interface design guidance: State of the art. Proceedings of the Human Factors Society, 1980.

Ramsey, H. R. and Atwood, M. E., Human Factors in Computer Systems: A Review of the Literature, Science Applications, Inc. Technical Report SAI-79-111-DEN, Englewood, Colorado, 1979. (NTIS No. ADA075679).

Rasmussen, J. Outlines of a hybrid model of the process plant operator. In Sheridan T.B., & Johannsen, G. (Eds.), Monitoring Behavior and Supervisory Control. New York: Plenum, 1976.

Reisner, P. Use of psychological experimentation as an aid to development of a query language. IEEE Transactions of Software Engineering, 1977, SE-3, 218-229..

Robinson, A.L. Perilous times for U.S. microcircuit makers. Science, 1980, 208, 582-586.

Scherr, A.L. An analysis of time-shared computer systems. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA: MAC-TR-18, 1965.

Sheil, B.A. Coping with complexity. In Kasschau, R., Lachman, R., & Laughery, K. (Eds.), Psychology and information processing technology: A view of the future, in press.

Siegel, A.I., & Wolf, J.J. Man-machine simulation models. New York: Wiley, 1969.

Stevens, A.L., & Collins, A.M. The goal structure of a Socratic tutor. Proceedings of Association for Computing Machinery National Conference. Seattle, WA, 1977.

Standardized Computer Resource Interface and Management Plan (SCRIMP), Center for Tactical Computer Systems (CENTACS), U.S. Army Communications Research and Development Command, Ft. Monmouth, NJ, December 1980.

Synectics Corporation, Development of Design Guidelines and Criteria for User/Operator Transactions with Battlefield Automated Systems, work in progress under ARI Contract No. MDA903-80-C-0094, Fairfax, Virginia, 1980.

Young, J.A. The future of electronics: Nine views from the top. Electronic Design, 1981, 29(1), 142, 143.

APPENDIX A
PLANNED CAPABILITIES OF FOUR OPERATIONAL FAMILIES
OF FUNCTIONS FOR THE WWMCCS INFORMATION SYSTEM

(from ASD WIS Report, pp. 30-33)

RESOURCES AND UNIT MONITORING

1. Monitor, determine, and display the status and readiness of all U.S. resources including active and reserve elements and appropriate non-U.S. forces and resources.

2. Schedule resource utilization.

3. Prepare and disseminate orders and mission instructions.

4. Plan operations, schedule missions, and monitor/display operations and mission results.

5. Identify and display unit, resources, facilities, communications, and weapon systems characteristics and capabilities.

6. Integrate and display data from systems external to WWMCCS such as environmental and intelligence data.

7. Monitor status of actions, critical events, situations assessment, rules of engagement, major end items of equipment, crisis situations, etc.

8. Determine net situation assessment.

CONVENTIONAL PLANNING AND EXECUTION

1. Generate and refine notional and actual force and resupply requirements and options.

2. Generate and refine force movement requirements.

3. Generate and refine notional and actual resupply and non-unit personnel movement requirements.

4. Merge, modify, and tailor force requirements and force lists from different plans.

5. Merge, modify, and tailor movement requirements from different plans.

6. Merge, modify, and integrate force, resupply, and non-unit personnel movement requirements.

7. Determine option/OPLAN force, logistic, and transportation availability and feasibility.

8. Match and select OPLAN notional force and resupply requirements with real-world forces and actual resources.

9. Develop, refine, coordinate, and disseminate appropriate movement tables, schedules, orders and plans.

10. Identify force, logistic personnel, and transportation shortfalls, limitations, and bottlenecks.

11. Rapidly reflow movement requirements and produce flow plans.

12. Conduct force, logistic, personnel, and transportation sensitivity analyses.

13. Merge movement requirements with channel traffic requirements.

14. Monitor the deployment of forces and material.

15. Monitor the movement of mobilized reserve forces from home station to mobilization station and the movement of non-deploying material within the CONUS.

16. Monitor the reception and onward movement of deploying forces and material within theaters of operation.

17. Identify location and status of airlift and sealift assets.

18. Provide near real-time crisis management information to include force, logistic, and movement data.

19. Provide information for the coordination of deployment routing, over-flight routes, and landing rights.

20. Provide information for the coordination of air refueling routes, requirements, timing and schedules.

21. Generate notional and actual logistic requirements in selected functional areas to include: (1) civil engineering, (2) non-nuclear ammunition, (3) Petroleum, Oil and Lubricants (POL), (4) medical, and (5) selected supply classifications.

22. Through interfaces with appropriate systems, provide for: (1) initial mobilization of reserve forces and the marshalling of logistic resources, (2) identification of the deployability status of deploying forces, and (3) identification of critical logistic resources.

23. Incorporate host-nation support.

24. Aggregate and summarize requirements and movement information.

25. Tailor force list to crisis operations.

26. Develop, modify, and evaluate the feasibility of potential courses of action.

NUCLEAR PLANNING AND EXECUTION

1. Monitor nuclear force status and weapons.

2. Plan development and analysis.

3. Attack, strike and damage assessment, and residual capability assessment.

4. Reconstitute and redirect forces.

5. Terminate hostilities and active operations.

6. Nuclear weapons information, characteristics, and capabilities.

7. Bomb-damage information.

8. Sorties timing, routing, and target assignment.
9. Provide target information.
10. Planning factors and system performance.
11. Revise or modify plans.
12. Monitor force generation.
13. Sortie launch.
14. Weapon strike.
15. Reconnaissance planning.
16. Plan execution.
17. Nuclear detonation information.
18. Re-targeting.

TACTICAL WARNING AND SPACE DEFENSE

1. Provide air and strategic missile warning.
2. Provide tactical missile warning.
3. Support the determination that an attack is in progress and an assessment of the nature of the attack.
4. Provide nuclear reconnaissance information for damage assessment.
5. Monitor status of nuclear capable forces.
6. Monitor environmental conditions.
7. Identify nuclear detonations.
8. Provide damage assessment.
9. Communication spot reports.
10. Monitor space defense force status and weapons.
11. Plan development and analysis.

12. Attack, strike, damage, and residual capability assessment.

13. Reconstitute and direct forces.

14. Terminate active operations.

15. Space defense weapon information, characteristics and capabilities.

16. Intercept effectiveness information.

17. Target ephemer is prediction, intercept point generation, intercept profile determination, and target assignment.

18. Provide target information.

19. Provide hostile weapon information.

20. Provide vulnerability information on friendly assets.

21. Provide information on available countermeasures to defeat hostile operations.

22. Planning factors and system performance.

23. Revise or modify plans.

24. Monitor force generation.

25. Sortie launch.

26. Weapon strike.

27. Space surveillance system configuration and tasking.

28. Plan execution.

29. Re-targeting.